Norwegian School of Economics
Bergen, Spring 2021

Common Ownership and Tax Avoidance

Vy Ngoc Khanh Nguyen & Sindre Lunner Nyberg

Supervisor: Maximilian Todtenhaupt

Master thesis in Accounting and Auditing (MRR)

NORWEGIAN SCHOOL OF ECONOMICS (NHH)

Table of Contents

Abstract **6**

Acknowledgement **7**

I. Introduction **8**

II. Literature review **8**

2.1 Determinants of tax avoidance 8

2.2 Ownership structure as a determinant 8

2.3 Framework for network ties studies 9

2.4 Board interlocks as peer group definition 10

2.5 Industries as peer group definition 10

2.6 Common institutional shareholder as peer group definition 10

2.7 How large shareholders might affect tax strategies 11

2.8 The role of industries in tax planning 12

III. Hypothesis development **13**

IV. Introduction to our model **15**

4.1 Panel data 15

4.2 Fixed effects with unbalanced panel 16

4.3 Caveat: the reflection problem 16

4.4 Caveat: firm fixed effects 19

V. Definition of variables **20**

5.1 Dependent variables 20

5.2 Independent variable 20

5.3 Control variables 21

5.3.1 Intangible assets 21

5.3.2 Research and Development 22

5.3.3 Missing Research and Development 23

5.3.4 Firm size 23

5.3.5 Value and stock return 23
5.3.6 Leverage ratio 23
5.3.7 Profitability 24
5.3.8 Loss Carryforward 24
5.3.9 Change in Loss Carryforward 24
5.3.10 Plan, Property and Equipment 25
5.3.11 Institutional ownership percentage 25
5.3.12 Foreign income or loss 25
5.3.13 Equity in earnings 25
5.3.14 Herfindahl-Hirschman Index 26
VI. Data processing 27
6.1 Data sources 27
6.1.1 Data source: Common ownership data 27
6.1.2 Data source: CRSP/Compustat merged 28
6.1.3 Processing shares in the common ownership data 29
6.2 Merging the datasets 29
6.3 The effective tax rates 31
6.4 Handling of missing values 33
6.4.1 Missing R&D 33
6.4.2 Other variables 34
VII. Sample construction 34
VIII. Empirical analysis 35
8.1 Empirical analysis preface 35
8.2 Summary statistics 35
8.3 ETR graphs 37
8.4 Correlation plots 39
8.5 Baseline regression 40

8.6 Differences in the peer ETR effect: peer firms connected by long versus short-term blockholders 42

8.7 Differences in the peer ETR effect: peer firms with long versus short-term blockholders 45

IX. Supplementary empirical analysis 46

9.1 Differences in the peer ETR effect: peer firms connected by the largest blockholder versus peer firms connected by other blockholders 46

X. Empirical analysis: robustness tests 48

10.1 Firm fixed effects 48

10.2 Peer firm restriction based on SIC 49

10.3 Event study on shocks in peer ETR's 51

10.4 Event study on shocks in ETR's on the investor-level 54

XI. Limitations and alternatives 57

XII. Conclusion 58

References 60

Tables 65

Appendixes 95

List of Figures, Graphs, Plots, Tables and Appendixes

Figure 1.1: (Coinciding fiscal year and calendar year) 30

Figure 1.2: (Coinciding fiscal year and calendar year) 30

Figure 2.1: (Peer ETR calculation) 31

Figure 2.2: (Peer ETR calculation) 32

Figure 2.3: (Peer ETR calculation) 32

Figure 2.4: (Peer ETR calculation) 33

Figure 2.5: (Peer ETR calculation) 33

Figure 3: (Sample selection procedure) 35

Graph 1: (Yearly ETR averages) 37

Graph 2: (Boxplot on ETR's) 38

Plot 1: (Correlation matrix for the GAAP ETR sample) 39

Plot 2: (Correlation matrix for the cash ETR sample) 40

Table 1.1: (Summary statistics) 65

Table 1.2: (SIC statistics of peer groups) 67

Table 2: (Baseline regression model) 67

Table 3: (Long-term vs. short-term blockholder in focal firms) 70

Table 4: (Long-term vs. short-term blockholder in peer firms) 72

Table 5: (Largest CIB regression) 75

Table 6: (Firm fixed effects regression model) 77

Table 7: (SIC restrictions) 79

Table 8: (Event study on shocks in peer ETR's) 86

Table 9: (Investor level event study) 91

Appendix A: (Variable definitions) 95

Appendix B: (Descriptive statistics of number of shocks in the sample from event study on peer ETR's) 96

Appendix C: (Missing values handling) 97

List of Abbreviations

TR	Thomson Reuters
WRDS	Wharton Research Data Services
SEC	Securities and Exchange Commission
CIK	Central Index Key assigned by the SEC for this investor
CUSIP	The identity of the holdings based on SEC's 13 (f) listings
rdate	Reporting date (end of quarter)
fdate	Filing date
ftype	The form name
CIB	Common institutional blockholder
ETR	Effective tax rate
GAAP	Generally Accepted Accounting Principles
SIC	Standard Industrial Classification
NoCeT	Norwegian Centre for Taxation
CRSP	Center for research in security prices

Abstract

There have been significant priorities among tax practitioners and policymakers about corporate tax planning. There are many ways firms can avoid tax. However, we know very little about how firms learn about different tax avoidance mechanisms. One of the crucial channels is common owners, specifically common institutional blockholders (CIB's), who potentially hold a momentous role in expediting the diffusion of tax avoidance knowledge across firms. Do firms engage in a similar level of tax avoidance if they share the same CIB's? We investigate this question using the *Common Ownership Data* and *Compustat/CRSP balance sheet data* via Wharton Research Data Services (WRDS). Our final sample results in 23,603 (23,015) observations from 1999 to 2016 using GAAP ETR (cash ETR) as tax avoidance measure, from the raw data of around a 48million observations of holding information.

Similar to a prior study by Cheng, Sun & Xie (2018), our empirical results support that firms follow their peers held by the same CIBs in making their tax avoidance strategies. We examine the causality of the peer effect on the focal firm. It is supplemented with other analyses using exogenous events, i.e., tax rate shock among peer firms. We acknowledge that companies operating in the same industry share the same firm-level characteristics; to proscribe this effect, our models look at peer firms with different SIC-code in relation to the focal firm. In addition, we conduct an event study on the investor level to observe blockholders adjust their portfolio weights when there is a shock in the focal firm's ETR. However, the effect is limited to using cash ETR only, with blockholder adjust their portfolio when there is a negative cash ETR event. We also conduct an analysis that differentiates between short- and long-term blockholders.

Our findings suggest that there is a positive interrelation between GAAP (Cash) ETR and PEER ETR. The SIC-code model indicates that investors are proved to drive the effect we see instead of latent industry-based characteristics. Furthermore, we find that short-term blockholders have more impact on the focal firm's tax adjustment. Overall, our finding supports the hypothesis that firms engage in similar levels of tax avoidance if common blockholders own them. Our results support the tax avoidance effect is not driven by endogeneity issues.

Further, our study aims to document a potentially vital channel for tax avoidance diffusion through CIBs. We hope that this thesis contributes to taxation and public finance by constructing and analysing the impact of common ownership through strategic tax avoidance mechanisms. It will also provide avenues to provide further research on this novel topic.

Acknowledgment

The thesis is written as a part of the Master of Science in Accounting and Auditing at the Norwegian School of Economics (NHH) and our major in Accounting and Auditing. The thesis is written under the supervision of the Norwegian Centre for Taxation (NoCeT).

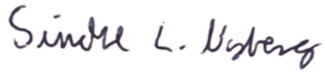
We are both grateful for and humbled by being selected to be a part of this research project. We were thankful for receiving good feedbacks from our supervisor from the Norwegian Center of Taxation. Besides, we are delighted to be selected for the grant for master theses in tax economics provided by the Norwegian Center of Taxation and the Norwegian Tax Administration.

We would like to express our sincere gratitude to our supervisor, Maximilian, for giving us helpful insights and great discussions through constructive meetings and guidance.

Finally, we offer our warmest thanks to our families, friends, and fellow students for their trust, continuous assistance, and support.

Oslo & Bergen, 28.05.2021

Vy Ngoc Khanh Nguyen



Sindre Lunner Nyberg

I. Introduction

Common ownership is described as where two firms are at least partially owned by the same investor, and this phenomenon is rising among competing publicly held U.S. firms. Nearly 25% of the U.S. stock market is now held by index funds, a rapid increase from 15% in the early 2000s. The world's three largest index funds (Blackrock, Vanguard, and Fidelity) are now holding a total of 18,17 trillion dollars under its management, including shares in many of the U.S.'s largest firms. The rise of common ownership has captured attention from academics and practitioners about the significant effect of this phenomenon on a firm's strategic choices (Easterbrook & Fischel, 1982; Hansen & Lott, 1996; Rubin, 2006), among which might be concerned about possible behaviour such as corporate tax planning.

In this paper, we try to answer how firms learn about different tax avoidance mechanisms, specifically via common owners. We examine whether firms engage in the same level of tax avoidance if the same institutional blockholders own them.

II. Literature Review

2.1 Determinants of tax avoidance

It is a major area of interest to examine determinants, magnitudes, and consequences of corporate tax avoidance (Hanlon & Heitzman, 2010). For example, Shackelford & Shevlin (2001) state that we know little about cross-sectional differences in the willingness of firms to avoid taxes and encourage accountants to participate in tax-related research. Companies, just as individuals, vary in their tax aggressiveness. Questions arise about the determinants of such tax aggressiveness. Related to this paper, ownership structure is the determinant of interest. Desai & Dharmapala (2009) is a study in which they used institutional ownership as a proxy to measure the quality of firm governance when they were to study how the quality of firm governance affects tax avoidance on firm values.

2.2 Ownership structure as a determinant

Several previous studies have also examined how ownership structure affects tax avoidance. Chen, Chen, Cheng, & Shevlin (2010) use ownership structure as a determinant to investigate tax avoidance. Specifically, they distinguish between family-owned companies and non-family-owned companies and find evidence that family-owned companies are less tax aggressive than non-family-owned companies. Another study uses a different variant of the ownership structure determinant and looks at how activist hedge funds affect tax avoidance (Cheng, Huang, Li, & Stanfield, 2012). Hedge fund activists buy into companies that perform poorly, hoping to change the company's firm value. Their finding is that targeted firms get an increased tax avoidance after interference from hedge fund activists. Furthermore, Khurana & Moser (2013) examines whether the level of institutional ownership with a long-term horizon is associated with the company's tax avoidance. They find evidence that there is less tax avoidance in firms held by long-term institutional owners.

In our studies, we also use ownership structure as a determinant to investigate the effect of tax avoidance. Our variant is based on Cheng, Sun, & Xie (2018) that we look at common institutional ownership and investigate whether there is knowledge sharing across firms with the same institutional owners. Hence, we study the peer effects of tax avoidance for firms who have the same institutional owners following Cheng, Sun, & Xie (2018). From what we have found, only the article mentioned above defines the peer group similarly as we do in our analysis.

Furthermore, we respond to Hanlon & Heitzman (2010), who mention that ownership structure might be an exciting determinant for examining the effect of tax avoidance.

2.3 Framework for network ties studies

There are many ways to study social network ties. Carpenter, Li & Jiang (2012) serve framework to guide scholars in conducting social network research in terms of research about interpersonal level or inter-organizational level. They propose whether to conduct social capital research or network development research. Social capital research means research that looks at network causes and predictors. In contrast, network development research is based on the network itself being the consequence and examining this formation and change. Following this framework, we are in our study in social capital research since we look at peer firms as

predictors for focal firms' ETR's. Furthermore, we are within the inter-organizational level to the extent that the institutional investors are organizations and not people.

2.4 Board interlocks as peer group definition

Other studies look at the effect of interpersonal linkages, such as the consequences on loan terms when people in the management have attended the same school or worked together with those applying for loans (Engelberg, Gao, & Parsons, 2012).

Gulati & Westphal (1999) find evidence that board interlocks can influence the formation of strategic alliances between companies, especially if there is no significant degree of independent control over the board and a large degree of CEO-board cooperation. Hence, board interlocks make an ideal setting to examine the impact of social ties on firm behaviour since the board itself makes the corporate decisions (Brown & Drake, 2014).

Brown & Drake (2014) examines tax avoidance when using board interlocks to define which companies are affiliated. The study looks explicitly at low-tax ties: low-tax companies in the bottom quintile based on industry-adjusted, long-run cash-effective tax rates. They argue that board ties to low-tax firms with operational and strategic similarities with focal firms should affect tax avoidance at the focal firm to a greater extent because the information shared between the companies is more relevant. There are better opportunities for focal firms to absorb and use tax avoidance opportunities in their business. They find operationally and strategically similar network ties significantly impacting focal firms' tax avoidance than dissimilar network ties. They use several different determinants on firm-level characteristics to define which companies are part of a peer group.

2.5 Industries as peer group definition

Another study examines companies mimicking product market leaders within the industry. It is an incentive for rival firms to mimic the product market leader's firm's performance, including tax avoidance, to remain competitive themselves (Kubick, Lynch, Mayberry, & Omer, 2015).

2.6 Common institutional shareholders as peer group definition

Unlike the main emphasis of previous studies in tax avoidance, our study differs to the extent that we look at peer groups where the companies are most often unrelated and thus have independent firm-level characteristics. Because we use a common investor base as a determinant to examine the peer effect of tax avoidance and the fewer investors build their portfolio with companies with the same level of tax avoidance ("ex-ante" preferences), the companies have few other relationships. For example, like Cheng, Sun & Xie (2018), few peer groups are in the same industry. 5.74% of the companies in our database on average share the same 4-digit SIC, 8.21%, 11.62%, and 19.83% for 3,2 of the firms share 3-digit, 2-digit, and 1-digit SICs, respectively.

Furthermore, as additional research, we investigate whether industry-level characteristics drive the results. We drop peer firms that share the same SIC as the focal firm. The reason for this is because, in contrast to Cheng, Sun & Xie (2018), we have a higher number of shared industry firms in our peer groups (see Table 1.2) and want to see if the results persist.

2.7 How large shareholders might affect tax strategies

A study by Cronqvist & Fahlenbrach (2009) examines how large shareholders affect corporate policies and firm performance. Among other things, they find that blockholders significantly influence investment, financial, and executive compensation policies. It indicates that blockholders have varying preferences, investments, and governance styles. Hence, the study shows that blockholders are also an important piece that affects firm performance differences. When it turns out that large institutional blockholders influence corporate behavior by influencing firm's decision-making, it will mean that these investors also have a golden opportunity to share their mindsets and preferences with the management of the companies (Cheng, Sun, & Xie, 2018). With this background in mind, our study uses the critical assumption that investors can influence the management of focal firms regarding tax avoidance policies. Thus, it is assumed that companies learn tax avoidance policies from other companies with the same investor base through, for example, social learning mechanisms. This mechanism is described by Cheng, Sun & Xie (2018) as investors "vote with their voice" Another possible mechanism may be that the firm's management itself tries to satisfy the blockholders. Therefore, the management might examine the blockholders' preferences by looking to peer firms which is a part of the blockholders' portfolio. This mechanism is described by Cheng, Sun & Xie (2018) as investors "vote with their feet". The effect simply

comes from the fact that investors are present and have an exit opportunity. Both mechanisms lead to companies learning from their peer firms when it comes to tax avoidance policies. Furthermore, Bird & Karolyi (2017) cf. Bird & Karolyi (2018) examines the effect of institutional ownership on tax avoidance. They find evidence that positive shocks to institutional ownership around Russell index reconstitutions lead to decreased effective tax rates. In addition, institutional ownership leads to reductions in ETR for companies with high ETR, while the effect is the opposite for low-ETR firms. This fact means that institutional owners contribute to pushing companies towards a common level of tax avoidance. Concerning our analysis, which we will return to later in the thesis, we include institutional ownership in percent as a control variable.

2.8 The role of industries in tax planning

Heitzman & Ogneva (2019) finds evidence that equity returns increase with the propensity for tax planning in a firm's industry. They developed a hypothesis that asks whether investors require compensation to invest in firms that compete in industries characterized by greater tax planning intensity. They approach the problem based on critical assertions that the firm's tax planning is heavily influenced by its industry, leading to investors expecting firms in industries with ample tax planning opportunities to adopt similar future tax strategies. Furthermore, an intensive tax-planning industry will increase the expectation of the firm's future exposure to non-diversifiable risk.

There is abundant evidence that industry plays a central role in the tax planning environment. For instance, at the legislative level, the Treasury's annual Tax Expenditures reports in detail "revenue losses attributable." The industry characteristics question whether there is an interaction between industry and firm tax planning beyond externalities. We all expect that firms operating in the industry with high scrutiny from the tax authority will have lower net cash flows from its tax planning, reducing the cost of capital benefit suggested by Shevlin, Lim, Lee & Goh (2016). In addition, high scrutiny drives the systematic risk of tax planning upwards if the tax authority resists a particular method or instrument of tax planning increases the correlation of the firm's cash tax savings with those other firms. There is a robust anecdotal and empirical support for the centrality of industry in firms' tax planning decisions (Heitzman & Ogneva, 2019).

III. Hypothesis development

Hanlon & Heitzman (2010) asks for more research on tax avoidance while using ownership structure as a determinant. While a stream of literature uses ownership structure as a determinant for tax avoidance research (Chen, Chen, Cheng, & Shevlin, 2010; Cheng, Huang, Li, & Stanfield, 2012; Khurana & Moser, 2013), only Cheng, Sun & Xie (2018) have examined the importance of a common institutional investor base. They find that common institutional investors have an impact on corporate tax avoidance.

Furthermore, Cheng, Sun & Xie (2018) argues for two distinct ways in which common institutional blockholders could influence the firm's management regarding tax avoidance.

Firstly, institutional investors could vote with "their feet" by selling their shares when they are dissatisfied with corporate performance (Parrino, Sias, & Starks, 2003). Thus, the firm's management could be encouraged to look towards other firm's owned by the same blockholders and adjust their tax avoidance strategies in line with these peer firms. This means CIB's have an indirect influence on firm's tax avoidance.

Secondly, there is also the possibility that institutional investors have a direct influence on corporate behaviour, thus "voting with voice." However, the literature is mixed regarding which institutional investors influence corporate decisions (Chen, Harford, & Li, 2007). Some institutional investors do indeed influence corporate behaviour while others do not.

However, Admati, Pfleiderer & Zechner (1994) argue that monitoring by large shareholders allows the shareholder to influence the firms following their interest through negotiation with management, proxy fights, and involvement in choosing members of the board. This means CIB's have a direct influence on firm's tax avoidance by pressuring management.

Since institutional investors have their preferences for tax avoidance, they are likely to serve as a bridge to facilitate the diffusion of tax avoidance information across firms in their portfolios (Cheng, Sun, & Xie, 2018).

Furthermore, Edmans (2009) argue that blockholders can impact governance either by "voting with voice" or "voting with feet." Thus, both mechanisms could have an impact on a firm's tax

avoidance, and we do not distinguish between these two mechanisms for our main research question.

This leads us to our main hypothesis:

H.1: *Firms engage in similar levels of tax avoidance if they are owned by common institutional blockholders.*

One question is when tax avoidance mimicking occurs. Is it the case that the companies continue to mimic tax avoidance to their CIB connected peer firms long after an investor has become a blockholder (ie owned 5% of shares outstanding), or is it the case that the mimicking is greatest in the first years, and gradually decreases?

Other studies in tax avoidance (Baskaran, 2015; Lopes da Fonseca, 2017) find that declining tax mimicking is the case. We want to investigate whether this can also apply in our institutional setting where we look at U.S. firms connected by CIB's.

Furthermore, Cheng, Sun & Xie (2018) argue that blockholders with long-term horizons should have a greater influence on corporate behaviour because of a more stable influence. Thus, long-term blockholder are more likely to get involved in decision making. They study the peer relation, where long-term relation is defined as peer connected by CIB's for more than one year in the five years.

In our analysis, we differ from this approach. We study the effect by how long the blockholder have been a blockholder in the firms and not the relation itself.

We also predict a portfolio weighting difference between short-term blockholders and long-term blockholders. Short-term blockholders have more recently bought shares, thus surpassing the 5% of shares outstanding mark. This does not mean long-term blockholders cannot trade shares while still being above the 5% of shares outstanding mark. Regardless, we expect more portfolio weight changes for blockholders classified as short-term.

We conduct two different analysis which is distinct in the following way. Firstly, we examine firm-year observations which contains both short-term and long-term blockholders. Then we

calculate peer ETR based on focal firms short-term blockholders other portfolio firms. Then we do the same for long-term blockholders in the focal firms. In this analysis it does not matter whether the blockholder is classified as short-term or long-term in the peer firms. Thus, we want to see whether the classification in the focal firms matter.

Secondly, we conduct another analysis. However, this time around we do not examine the classification at the focal firms. Instead, we examine whether the focal firms react upon peer firms ETR's based upon the CIB's are classified as a long-term blockholder or short-term blockholder in the peer firms. Thus, we want to see whether the classification in peer firms matter.

With how the portfolio weighing is calculated in mind (see chapter 6.3), and the findings from Baskaran (2015) and Lopes da Fonseca (2017) we develop two supplementary hypotheses:

H.2.1: *Tax avoidance mimicking towards peer firms are stronger when peer firms are connected by short-term blockholders in the focal firms.*

H.2.2: *Tax avoidance mimicking towards peer firms are stronger when the connected blockholders are short-term blockholders in the peer firms.*

IV. Introduction to our model

4.1 Panel data

The variable α_i captures all unobserved, time-constant factors that affect y_{it} . a_i is called the unobserved effect. α_i is often referred to as unobserved heterogeneity (or individual heterogeneity). In our models, individual heterogeneity refers mainly to firm heterogeneity, industry heterogeneity, and year heterogeneity. The error u_i is often called idiosyncratic error or time-varying error because it represents unobserved factors that change over time and affect $y_{it.}$ These are very much like the errors in a straight time series regression equation. A simplification of our models is:

$$ETR_{it} = \beta_0 + \beta_1 PEER\_ETR_{it} + \beta_2 Control\ variables_{it} + \alpha_i + \mu_i$$

With t = 1, 2, 3….and so on

Since i denotes different firms, we call α_i an unobserved firm effect: it represents all factors that can affect the focal firm tax rate, which does not change over time. We have included firm characteristics in α_i. However, as explained earlier, fixed effects remove some relevant variation since we want to capture strategic reaction, which is constant over time (i.e., inherent managerial characteristics). Hence, we focus on industry fixed effects and not firm-fixed effects in most of our analysis.

4.2 Fixed Effects with Unbalanced Panel

Our database contains missing years on individual firms since their shareholding information and change from year to year. Investment time length also varies. This results from the fact that some firms do not have CIBs across all years in the sample. More importantly, is the fact that firms' effective tax rate also needs to be in the 0 to 100% range even to be considered in the peer effective tax rate calculation, and this varies across years.

Some investors hold shares in a shorter window, while others hold a more extended time period, which makes t variates. Another common scenario is when an investor liquidates his shares in one firm, making his ownership leaves the sample, so-called attrition. A problem arises when attrition is correlated with the idiosyncratic error – those observed factors that change over time and affect effective tax rate – then the resulting sample section problem can cause biased estimators (Baltagi, Song, & Koh, 2003). We also have a non-random sample in subsequent periods. However, we are aware of these inherent characteristics of our data set, and this limitation is considered in our model.

Provided that some i is not correlated with the characteristic errors, u_{it} , an incomplete panel data withhold. Even so, our model with firm fixed effects does allow attrition to be correlated with a_i, the unobserved effect. It aligns with the idea of allowing some units to be dropped out of the initial sampling, and a_i captures this. We also conduct fixed firm effects analysis to control for unobserved factors that can affect the focal's firm effective tax rate and be correlated with the fraction (Wooldridge, 2012) as a robustness test and in event studies.

4.3 Caveat: the reflection problem

Manski C. F. (1993) sheds light on an econometric challenge that makes it difficult to isolate the effect of the company's response to peer companies' tax planning strategies. Armstrong, Glaeser & Kepler (2019) points out the problem in a corporate tax planning setting and argue that the reflection problem comes from the fact that peer companies, specifical companies in the same industry, often have the same background characteristics and equal firm-level characteristics. This problem affects both companies' tax planning.

In our analyses in section 10.3, we only look at peer companies with different SIC codes in relation to the focal firm. In this way, we remove, to some extent, the problem within companies in the same industry that often have the same firm-level characteristics. We then see if focal firms follow their peer firms' tax planning strategies, where focal firms and peer firms are not within a common industry.

Manski C. F. (1993) illustrates "the reflection problem" by a trivial example. His model explains that the problem arises when a researcher observes the distribution of the behaviour in a group and tries to deduce whether the average behaviour in group A affects group B, where group A and B are the two subsets of the entire population. Imagine an analogy to reflection from looking in a mirror when you see someone make a hand movement. Without understanding what is going on, it becomes difficult to say whether the reflection in a mirror when you see someone makes a hand movement. Without knowing what is going on, it becomes difficult to say whether the reflection in the mirror affects the hand movement or whether it is the movement that affects the reflection in the mirror. To put it bluntly, an animal will not necessarily understand whether it reflects the mirror or the hand movement which is a triggering factor.

Hence, we use a lead-lag model in our analyses instead of seeing how focal firms follow their peers in the same year. In addition to the problem, new tax rules can affect both focal firms and peer firms in the same year.

Manski C. F. (1993) examines the reflection problem in linear models and categorizes three different determinants that affect the analysis result.

1. Firstly, we have endogenous effects, where an individual's propensity to behave in a given way varies with the group's behaviour.

2. Secondly, we have exogenous effects. An individual tendency to behave in a given way stems from the fact that individuals have the same individual or institutional characteristics as groups do.

3. Thirdly, we have correlated effects, where an individual firm's tendency to behave in a given way varies depending on the exogenous characteristics of the group.

Armstrong, Glaeser & Kepler (2019) describes these problems in its specific corporate tax planning setting, where peer firms are defined as firms in the same industry. Exogenous effects (2) can be caused by firms operating in the same industry being exposed to the same levels of tax planning due to the industry's environment. For example, companies within the same industry have the same opportunities for tax planning due to industry-specific tax rules, and these companies have to adapt to new tax strategies due to equal competitiveness in the peer group to survive. Our model with specifications ranging from 4-digit SIC code to 2-digit tries to solve this problem.

As for correlated effects (3), firms in the same industry may face the same level of tax planning because the firm-level characteristics of the firms are correlated. It can be similar risk profiles and capital structures for companies in the same industry. It is also a problem with the way we define peer firms. It is conceivable that investors choose companies for their portfolios with similar risk profiles and other firm-level characteristics.

Armstrong, Glaeser & Kepler (2019) describes the endogenous effects (1) as "outcome-on-outcome"" effects. The effects are beyond what is captured by the (2) exogenous and (3) correlated effect components. This direct effect represents how a company responds directly to a strategic tax change within their peer firms. We can interpret this effect component as a direct "strategic reaction."

One possible problem is that it is not easy to separate and isolate the effects from (1), (2), and (3), respectively. And in that case, we have an identification problem. Firstly, we have an endogeneity problem with possible omitted correlated effects (3). In our analysis, this problem may be due to investors building their portfolios with companies with the same firm-level characteristics that result in an equally effective tax rate. Secondly, we have the

endogeneity problem, making it challenging to separate direct strategic direction from (1) and (2). Even if one does not omit correlated products due to the simultaneous nature of strategic reaction (Armstrong, Glaeser, & Kepler, 2019). Focal firm number 1 responds to its peers. Still, this firm can also be a peer firm for another company; in this case, focal firm number 2 (where focal firm 1 is the peer company of focal firm two and vice versa), resulting in collinearity. We have a 11,1% (7,5%) correlation between focal GAAP ETR (focal cash ETR) in year *t* and peer GAAP ETR (peer cash ETR) in year *t-1* (see correlation plots in section 8.4).

To solve this endogeneity problem, we introduce a tax avoidance shock event for the peer companies, exogenous for the focal firm. Section 10.3 shows the results of the event study model.

4.4 Caveat: firm fixed effects

Armstrong, Glaeser & Kepler (2019) argues that strategic reaction contains a cross-sectional component that only varies among companies but is, on the other hand, fixed over time. When using firm fixed effects, we will remove some of the relevant variation required by answering the central question: Do firms engage in similar levels of tax avoidance if the same investors own them? Thus, when using firm fixed effects, we do not capture all relevant variation.

Within transformation removes all time-invariant variation. However, strategic reactions are captured in such variance, and we do have the problem with applying firm fixed effects. Some of the variety we are looking for disappears. Hence, we follow previous studies in the field and focus on regressions with industry fixed effects to remove variation between industries instead of between companies (Cheng, Sun, & Xie, 2018). More specifically, we use a 2-digit SIC as a basis for industry fixed effects. This approach allows us to control for unobservable heterogeneity between industrial groups while at the same time maintaining most of the relevant variation between companies.

In addition to the concern that fixed effects fail to understand the cross-sectional variation of strategic reaction, Leary & Roberts (2014) also mention that fixed effects should be used cautiously due to aggravation of measurement errors.

We still show that our results are robust to firm fixed effects and the use of this does not affect our baseline regression by a lot.

V. Definition of variables

5.1 Dependent variable

In this section we introduce the measures of tax avoidance.

One measurement used in this paper to calculate tax avoidance is the effective tax rate (ETR). Effective tax rate calculation is a valuable metric for benchmarking the tax paid between two or more entities. Companies pay tax on their gross income, and after-tax deductions sit firms with net income. In addition, investors actively use an effective tax rate as a profitability indicator of a company. Hence, we include ETR in COMPUSTAT actively in our model to determine the level of tax avoidance and decision to choose investments by investors.
We annual GAAP effective tax rate as a measure of tax avoidance. We follow the formula given by De Simone, Nickerson, Seidman & Stomberg (2020) for calculating GAAP ETR:

$$GAAP\ ETR = \frac{Total\ Tax\ Expense}{Pretax\ Income}$$

Furthermore, we also use cash effective tax rate as another measure of tax avoidance. We follow the cash ETR formula given by De Simone, Nickerson, Seidman & Stomberg (2020):

$$Cash\ ETR = \frac{Income\ Taxes\ Paid}{Pretax\ Income}$$

5.2 Independent variable

In this section we introduce the independent variable of interest, namely the peer ETR's.

According to Cheng, Sun & Xie (2018), a novel channel of tax avoidance diffusion is measured by connections between firms sharing common institutional blockholders (CIBs). We define firms as peer firms when at least one institutional blockholder (which owns a minimum of 5% of the currently outstanding shares) is shared between these firms.

A blockholder is described as the owner of a significant block of a firm shares and bonds who can influence the company with their voting rights. Hence, our peer ETR is based upon firms connected through CIBs.

As for the calculation of peer ETR, firstly, peer firms are identified through common institutional blockholders with focal firms. Then we calculate a portfolio weighted ETR based on all the peer firms, excluding the focal firms. This is done iteratively. Consider firm A, B, and C as a peer group defined by one shared CIB. Peer ETR for firm A is calculated as the ETR of firm B and C. Thus, firm B and C is the portfolio excluding firm A, the focal firm.

In the next iteration, firm B is considered as the focal firm, and firm A and C represents the portfolio excluding firm B. Hence, peer ETR for firm B is calculated based upon firm A and C's ETR while weighting those two ETRs by the respective market capitalization ratio between firm A and C.

In all tables using cash (GAAP) ETR as a dependent variable, peer ETR is calculated based upon cash (GAAP) ETR.

See the chapter 6.3 for a trivial example of the peer ETR calculation.

5.3 Control variables

Our selection of independent variables are consistent with the existing literature (Cheng, Sun, & Xie, 2018).We control for intangible assets (*INTAN*), institutional investors (*INST*), leverage ratio (*LEV)*, market-to-book value (*MB*), loss carryforward (*NOL*), change in loss carryforward (*DIFFNOL*), net plant, property and equipment (*PPE*), profitability (*ROA*), logarithm of total market capitalization (*LOG_MKTCAP*), foreign income and loss (*FIN_C*), equity income in earnings, deflated by lagged total assets (*EQUINC*), research and development (*R&D*), industry-average research and development (*R&D_Blank*), Herfindahl-Hirshman Index (*HHI*).

5.3.1 Intangible assets

$$Intangible\ Assets\ \ Ratio\ (INTAN) = \frac{Total\ Intangible\ Assets\ (intan)}{Total\ Assets\ (at)}$$

It is traditionally challenging to evaluate the actual value of intangible assets since all expenses of creating intangible assets by companies cannot account for on the balance sheet, but only possible if the firm acquires these types of assets from others. Usually, the purchase price is much higher than the recorded book value of intangible assets, presenting the premium for goodwill. Neubig & Wunsch-Vincent (2018) examines a typical method for profit shifting among firms is the strategic location of intangibles. They find profit shifting is higher for entities with significant intangible assets because mispricing is more accessible and influential. Karkinsky & Riedel (2012) show low tax rates increase the profitability that a firm applies for a patent in a low-tax rate location. Beer & Loeprick (2015) support the hypothesis that profit-shifting responsiveness is higher for subsidiaries with higher ratios of intangible assets to total assets.

5.3.2 Research and Development (R&D)

$$R\_D\ = \frac{Research\ and\ Development\ Expense\ (XRD)}{Total\ Assets\ (AT)}$$

Firms' R&D expenditure tax deductibles and historically generate further tax credits. Mangers are motivated to classify corporate expense as R&D to reduce tax liabilities. This complex tax practice is labelled as "strategic R&D classification," thus places limitations on regulating firms' use of R&D tax credits. R&D tax credits create a permanent difference between a book and taxable income, resulting in a corresponding decrease in tax expense on the income statement (Laplante, Skaife, Swenson, & Wangerin, 2019). Many studies found high levels of responsiveness to profit shifting, particularly for R&D-intensive companies (Neubig & Wunsch-Vincent, 2018).

5.3.3 Missing Research and Development

We constructed a dummy variable for dealing with missing R&D observations. We combine the use of dummy variable for missing R&D with replacing missing values with the industry average of reported R&D (Koh & Reeb, 2015).

5.3.4 Firm size (MKTCap/LogMktcap)

$$Mktcap = PriceClose - Annual - Fiscal(PRCC\_F)\ x\ Common\ Shares\ Outstanding\ (CSHO)$$ [1]

There is an absence of scale and complexity in small firms than large firms. Smaller firms have less diversified tax strategies and are more exposed to changes in tax practices. Unlike their larger counterparts, they are less willing to invest in political influence to protect their strategies. Hence, they are more exposed to tax policy changes. Contrarily, large multinational firms have more incentives and conduct more efforts to shift income, namely shifting income through an alternation of transfer price assigned to international trade with affiliates (Clausing, 2009).

5.3.5 Value and stock return MB (Market-to-book value)

$$Market - to - book\ value\ (MB) = \frac{Price\ Close\ -\ Annual\ -\ Fiscal\ (PRCC\_F)\ x\ Common\ Shares\ Outstanding\ (CSHO)}{Total\ Common/Ordinary\ Equity\ (CEQ)}$$

5.3.6 Leverage ratio (LEV)

$$\text{Leverage Ratio} = \frac{Total\ Long-term\ Debt\ (DLTT)}{Total\ Assets\ (TA)}$$

Debt Tax Shield = Tax Rate * Interest Rate * Debt

Even though leverage ratio has not always been included with other measures in proxies for tax planning, its relevance as a measurement of tax planning is indisputable. Debt gives access to interest tax deduction valuable to the firms (Kaplan & Strömberg, 2009). By utilizing the Leverage Ratio as a proxy for tax planning, we try to capture how firms create value by generating debt tax shields. Debt tax shields are known as reducing income taxes that result from taking an allowable deduction from taxable income. The American tax system

[1] We use the natural logarithm of *Mktcap* as our firm size control variable.

favours debt over equity, as firms do not get tax deductions on dividends and thus not on equity. The strategy explains the incentives for firms to lever up to generate interest tax deductions (Roald & Roti, 2015).

5.3.7 Profitability measured by return on assets (ROA)

$$ROA = \frac{Pretax\ Income\ (PI)}{Total\ Assets\ (AT)}$$

Return on assets (ROA) is usually used as a benchmark to analyse how a company's management efficiently uses its assets to generate earnings. The higher the *ROA*, the better. High profitability also means higher tax to be paid. Total assets imply company sizes, where the company size has also been used to indicate tax avoidance (Kartikaningdyah, 2019). Gunaasih (2021) confirms that profitability has a positive effect on tax avoidance through empirical research.

5.3.8 Loss Carryforward (NOL)

A loss carryforward is an accounting technique used to reduce the total amount of tax the company owes the government by applying the current year's net operating loss to the future year's net income. Suppose a company experiences a negative net operating income (NOI) in year one but positive NOI in subsequent years. In that case, it can reduce the firm's future profits using the NOL carryforward to record some or all the losses from the first year in the following years. This method is used to spread a current net operating loss over subsequent years' NOI. Subsequently, NOL carryforwards are recorded as assets on the company's balance sheet. It gives the company an economic advantage in future tax liability tax savings and is especially valuable during an economic downturn. Streitferdt (2013) prove firms with tax loss carryforward have a higher value than comparable companies without unused tax credits. Loss carryforwards can impact firm value significantly (Betker, 1995). As a result of a direct relationship between the amount of fewer carryforwards and the amount of tax paid, *NOL* is included.

5.3.9 Change in Loss Carryforward (ΔNOL)

We observe change in loss carryforward in firms by setting $\Delta NOL = NOL_t - NOL_{t-1}$. If the change is positive, ΔNOL will have a dummy variable equal to 1 or otherwise equal to 0.

5.3.10 Net Plan, Property and Equipment (PPE)

$$PPE = \frac{Total\ Net\ Property, Plant\ and\ Equipment\ (PPENT)}{Total\ Assets\ (TA)}$$

5.3.11 Institutional ownership percentage (Inst)

Leipälä (2017) finds a shred of slight evidence that supports the hypothesis that firms with more institutional investors exhibit higher levels of nonconforming tax avoidance. A small association between institutional ownership and conforming tax avoidance is also confirmed. There is a positive association between institutional investors and corporate tax avoidance since they bring tax planning knowledge to make firms' tax planning more effective (Chen, Huang, Li, & Shevlin, 2019; Khan, Srinivasan, & Tan, 2017). On the other hand, a study proposed by Grossman & Hart (1980) concludes that institutional investors value short-term over long-term, which primarily may lead to higher tax avoidance. But this effect is contradicted by the free-rider problem that firms with multiple institutional blockholders have, which may lead to lower tax avoidance.

5.3.12 Foreign income or loss (F_INC)

$$\text{F\_INC} = \frac{Pretax\ Income\ Foreign\ (PIFO)}{Totel\ Assets\ (AT)}$$

International affiliates with foreign operations arguably have more tax planning opportunities through shifting their income to foreign subsidiaries than purely domestic firms. Payoffs from tax strategies are more exposed to changes in currency exchange rates, foreign tax policies, trade policies, and among other things. Multinationals are also less exposed to domestic tax policies if they operate in multiple countries, which leads to possibly higher or lower than the tax rate if they had operated purely in the U.S.

5.3.13 Equity income in earnings, deflated by total assets (EQINC)

$$EQINC = \frac{Equity\ in\ Earnings - Unconsolidated\ Subsidiaries\ (ESUB)}{Total\ assets\ (AT)}$$

Equity income is referred to income from stock dividends. Investors are rewarded for earning shares in a company. Dividend-paying companies can be large, well-established companies with mature revenue and earnings. Most of them have a well-established scheme for targeted annual dividend pay-out rates. Paying dividends usually sends a clear, powerful message about a company's performance, showing a solid demonstration of financial strength. Nonetheless, many quickly expanding U.S. technology tech firms refuse to pay dividends since they use their withhold cashback into operations during pivotal growth stages. The reasons could be that not paying dividends may be more beneficial to investors from a tax perspective. Karjalainen, Kasanen, Kinnunen & Niskanen (2020) examine tax considerations as determinants of dividends decisions and earnings management. There is a strong link between firms' taxes and reported net earnings in most European countries.

5.3.14 Herfindahl-Hirschman Index (HHI)

$$H = \sum_{i=1}^{N} s_i^2$$

where $H \epsilon (\frac{1}{N}, 1)$

$$\begin{cases} H < 0.01\ (or\ 100)\ indicates\ highly\ competitive\ industry \\ H < 0.25\ (or\ 1\ 500)\ indicates\ an\ unconcentrated\ industry \\ 0.15 < H < 0.25\ (or\ 1\ 500\ to\ 2\ 500)\ indicates\ moderate\ concentration \\ H > 0.25\ (above\ 2\ 500)\ indicates\ high\ concentration^2 \end{cases}$$

The term Herfindahl-Hirschman Index is an acclaimed measure used widely in competition law to gauge the market concentration of specific industries. It measures the size of firms and indicates competition among them. It is defined as the sum of squares of firms' market shares

[2] "Horisontal Merger Guidelines" (08/19/2010)

within industries, usually limited to 50 largest firms, where the market shares are expressed as fractions:

VI. Data processing

6.1 Data sources

6.1.1 Data source: Common ownership data

We use “Common Ownership Data” by Michael Sinkinson.

The data set contains investors and holding information of the S&P 500 index firms from 1999 to 2017, which 1999 marks the beginning year of compulsory electronic information filing. Although this data set has some improvements from Thomson Reuter’s known limitations, there is no guarantee that the data set is flawless, since the source itself can contain incomplete and wrong information.

There have been found a number of issues with the Thomas Reuters (TR) “S34” used by many researchers in their studies about the common ownership hypothesis, specifically a lack of records of Blackrock Holdings and discrepancies when accessing data for constituent firms of the S&P 500 Index. The new dataset is set out separately of 13(f) holdings from the source documents, are all public and available electronically from the Securities and Exchange Commissions (SEC) websites. Wharton Research Data Services (WRDS) has worked extensively to improve the database. They first gathered all 13(f) fillings from 1999-2017, the corpus is over 318,000 filling and occupies ~25GB of space if unzipped. Firms of interest is holdings of all public firms with a market capitalization of at least $10M. They extracted the filling date, reporting date, and reporting entity (Central Index Key, or CIK, and CIKNAME) from the header of the file. The approach is they look for any lines that contains a CUSIP code that they were interested in, and then they attempt to determine the “number of shares” field and the “value” field. With the help of the downloaded stock price data form CRSP, the filling date and value (price*share) can be validated. Any derivative holdings descripting by phrases such as OPT, CALL, PUT, WARR are removed from the data. The final resulting

dataset has around 48M reported holdings (CIK-CUSIP) for all 76 quarters and between 4,000 and 7,000 CUSIPs with between 1,000 and 4,000 investors per quarter. The fields are[3]:

- CIK: the central index key assigned by the SEC for this investor
- CUSIP: the identity of the holdings
- Shares: the number of shares reportedly held.
- Rdate: reporting date (end of quarter). 8 digit, YYYYMMDD
- Fdate: filling date. 8 digit, YYYYMMDD
- Ftype: the form name

They did not consolidate separate or possibly related Black Rock entities. They also dropped any CUSIP-rdate observation where any investor in that CUSIP reports owning greater than 50% of shares outstanding, or the one where greater than 120% of shares outstanding are reported to be held by 13(f) investors.

There are also profit weight values (i.e. \kappa) for all firms in the sample. The fields are:

- CUSIP_FROM: beginning date of the holding
- CUSIP_TO: ending date of the holding
- KAPPA: profit weight values of the holding
- Quarter: quarter identification of the holding

6.1.2 Data source: CRSP/Compustat merged

W use the CRSP/Compustat merged database to extract the financial data. All our variables originate from the CRSP/Compustat merged fundamental annual database with two exceptions. Firstly, for identifying a blockholder, defined as an institutional investor who holds at least 5% of the current shares outstanding in at least one quarter within a specific year, we use the CRSP/Compustat merged fundamental quarterly database. Secondly, when calculating institutional ownership percentage as a control variable, we also make use of the quarterly current shares outstanding variable (*cshoq*) from the CRSP/Compustat merged fundamental quarterly database.

[3] https://sites.google.com/view/msinkinson/research/common-ownership-data

The databases are available at the website of Wharton university of Pennsylvania [4].

6.1.3 Processing shares in the common ownership data

We begin with the Common Ownership data set which comprises of three variables:

- CIK: unique code for each unique investor
- CUSIP: shareholding information for each unique firm
- Rdate: shareholding reporting date
- Shares

The data set is comprised of quarter data since shareholding registration must be conducted quarterly, however, the Compustat/CRSP merged fundamentals quarterly did not contain variables to calculate all our control variables. Therefore, we decided to eliminate quarter data and retain annual data within the Common Ownership data and merge this with the Compustat/CRSP merged fundamentals annual database. We then group the data after each investor, shareholding and registration date. Consequently, we calculate the number of shares each investor holds as the summation of quarterly registered shares within each year and divide this by four. Thus, this calculation of shares from quarterly values to annual values is used in the portfolio weighting together with annual CRSP/Compustat data.

6.2 Merging the datasets

In order to manage to merge “Common Ownership” data set with COMPUSTAT/CRSP-merged data, we make a change for the CUSIP variable. CUSIP-variable in the Common Ownership data set has a string of 8 characters, while it has a string of 9 characters in the COMPUSTAT data set. The ninth digit in COMPUSTAT data is an automatically generated control digit to avoid mistakes in creating CUSIPs. We, therefore, choose to exclude this last digit, which makes the CUSIP-variable from the two data sets comparable for matching. This approach created a few duplicates in which multiple CUSIPs had different 9-digit strings but the same 8-digit strings. We excluded those from the sample.

[4] https://wrds-www.wharton.upenn.edu/pages/get-data/center-research-security-prices-crsp/annual-update/crspcompustat-merged/

Furthermore, we exclude firms from regulated and financial industries (those with SIC codes between 4900:4999 and 6000:6999).

A very crucial point we have considered in our analysis is the urge to differentiate between fiscal year and calendar year. Fiscal year is based upon which calendar year most of the fiscal months are within.

If the variable *fyr (Fiscal year-end Month)* is equal to or exceeds 6, then the calendar year and fiscal year will be the same. Otherwise, we add one year to the financial year, so that calendar and fiscal year coincide.

Figure 1.1: Coinciding fiscal year and calendar year

Example when adding one year to the fiscal year: Most fiscal months are within 2004 and fiscal year is 2004. However, end of balance date is in the 2005 calendar year; thus, we do need to adjust.

Calendar year	Fiscal year	Months	Coincided fiscal and calendar year
2004		May (5)	
2004		June (6)	
2004		July (7)	
2004		August (8)	
2004		September (9)	
2004		October (10)	
2004		November (11)	
2004		December (12)	
2005		January (1)	
2005		February (2)	
2005		March (3)	
2005	2004	April (4)	2005

Source: own illustration

Figure 1.2: Coinciding fiscal year and calendar year

Example when not adding one year to the fiscal year: Most fiscal months are within 2005 and fiscal year is 2005. End of balance date is in the 2005 calendar year; thus, we do not need to adjust.

Calendar year	Fiscal year	Month	Coincided fiscal and calendar year
2004		August (8)	
2004		September (9)	
2004		October (10)	
2004		November (11)	
2004		December (12)	
2005		January (1)	
2005		February (2)	
2005		March (3)	
2005		April (4)	
2005		May (5)	
2005		June (6)	
2005	2005	July (7)	2005

Source: own illustration

6.3 The effective tax rates:

After calculating the effective GAAP tax rate as total income taxes divided by pre-tax income, we filter out data with values larger than one and smaller than 0, hence GAAP ETR ∈ [0,1]. This approach is unlike De Simone, Nickerson, Seidman & Stomberg (2020) who winsorizes ETRs to 0 and 1. We also filter out firms with negative or zero pre-tax income. The same approach is used when calculating cash ETR, except we use income taxes paid as the numerator.

Peer-effective tax rate (Peer_etr) is based on CIB information for each firm per calendar year; we observe whether specific blockholders are also blockholders of other firms. We aggregate each firm's total market capitalization in the subset where the firm is excluded and then multiply with the firm's ETR. After that, we calculate the specific blockholder's total market capitalization in their portfolio, excluding the firm we are observing. Peer_ETR is weighted and summed up.

Illustration: Considering firm A with the following subset

Figure 2.1: Peer ETR calculation

Investor (CIK)	Firm (CUSIP)	Number of shares	PRCC	ETR

1	A	100	10	0.3
2	A	100	10	0.3
3	A	100	10	0.3

The table below illustrates the subset for investor one excluding firm A

For CIB 1, can we observe all other firms which also have the same CIB. The illustration below shows how firm B's, C's, and D's have CIB 1, and the tax rate is determined. Peer effective tax rate is weighted after each firm's market capitalization on the whole investor's portfolio total market capitalization, excluding firm A

Figure 2.2: Peer ETR calculation

CIK (1)	CUSIP (2)	Shares (3)	PRCC (4)	ETR (5)	MKTCAP (6)	Total MKTCAP (7)	(6)/(7) = (8)	Weighted Peer_ETR (9)
1	B	80	7	0.33	560	1955	0.2864	0.094526854
1	C	90	6	0.34	540	1955	0.2762	0.093913043
1	D	95	9	0.35	855	1955	0.4373	0.153060905
						Sum Peer_ETR		0.341508951

Figure 2.3: Peer ETR calculation

CIK	CUSIP	Shares	PRCC	ETR	Peer_ETR (no average)
1	A	100	10	0.3	0.3415
2	A	100	10	0.3	
3	A	100	10	0.3	

The same procedure is done for CIB 2 and 3, and their inputs are enumerated in firm A in the same way. The final Peer_ETR for firm A is thence a weighted product of Peer_ETR from all the inputs of different CIBs. Let us assume that input to Peer_ETR from investor 2 and 3 is

respectively 0.3243 and 0.314,3, and they have an equal number of shares in firm A. We calculate the shares-weighted Peer_ETR accordingly for firm A.

Figure 2.4: Peer ETR calculation

CIK	CUSIP	Shares	PRCC	ETR	Peer_ETR
1	A	100	10	0.3	0.3415 * (100/300)
2	A	100	10	0.3	0.3243 * (100/300)
3	A	100	10	0.3	0.3143 * (100/300)
		=300	Sum Peer_ETR		0.3267

The result for firm A is amounted to Peer_ETR of 0.3267 in one particular year.
Figure 2.5: (Peer ETR calculation)

CUSIP	ETR	Peer_ETR
A	0.3	0.3267

Parallel computing packages in R made this for-loop possible in a reasonable amount of run-time. The concept of parallel computing is to carry out many calculations simultaneously and improve calculating capacity (Schmidberger, et al., 2009).

6.4 Handling of missing values

For all details about missing values handling, see appendix C.

6.4.1 Missing R&D

Koh & Reeb (2015) shows a perusal of subsample of 3000+ NYSE-listed firms in their samples shows that a substantial number fail to provide any information regarding their R&D efforts. This accounts for 57.9% of firms who fail to provide any information regarding their corporate R&D efforts, and 12.4% report zero R&D.

In the study by Koh & Reeb (2015), corporate R&D disclosure decision is investigated to be affected influentially by managers' nature discretion decision. Their study encounters non-reporting R&D firms file over 14 times as firms that report zero R&D expenditure, and more than 10% of COMPUSTAT firms display evidence of engaging in innovation and activities. Hence, indicating the interpreting missing R&D as zero R&D activities in firms can lead to substantive bias in our empirical test. In line with Koh and Reeb's Monte Carlo simulations recommendation, we denote a blank dummy for missing R&Ds. Our panel data benefit from replacing missing R&D values with the average industry R&D and a dummy variable to denote missing values replaced by the industry average (Koh & Reeb, 2015).

There is a total of about 8,000 missing R&D observations in our net samples, and these are replaced with the industry average R&D. Some industries do not have any firms with non-missing R&D values, and as a result, some observations have not been replaced with the industry average and is set to 0, aligned with the approach posted by various studies in Strategic Management (Koh & Reeb, 2015).

6.4.2 Other variables

We also replace missing pretax income(foregin) with pretax income minus pretax income(domestic) (Dyreng & Lindsey, 2009). Missing intangible assets is set to 0 (Peters & Taylor, 2017). Rest is explained in appenix C.

VII. Sample construction

Our procedure begins with all listed firms in COMPUSTAT data for fiscal years from 1999 to through 2017 by WRDS. We set up some requirements for observations to avoid selection bias, for which we have the necessary data to compute our primary tax planning measures and control variables. Firms in the COMPUSTAT dataset have to be incorporated in the U.S. (fic = "USA") from 1999 to 2017. This means excluding foreign corporations subject to the resident country with different tax laws than U.S. tax laws. We exclude firms in the utility industry (SIC: 4900-4999) and financial industry (SIC: 6000-6999). The utility sector is characterized by close customer-supplier relationships, enabling principal customers and dependent suppliers to affect a firm's tax avoidance strategy (Cen, Maydew, Zhang, & Zuo, 2017). The financial industry is also excluded since it is placed under the special corporate

franchise tax and complicated regulations. Like Armstrong, Glaeser & Kepler (2019) and Dyreng, Hanlon, Maydew & Thornock (2017), we exclude firm-years with negative pretax income. We winsorize our control variable with 1st and 99th percentiles. We also set our dependent variable so that ETR's lies in the unit interval between zero and one. We lag all firm-, year- and industry-level control variables by one year. Our final samples consist of about 23 thousand firm-year observations.

Figure 3: Sample selection procedure

	Number of observations (GAAP ETR procedure)	Number of observations (Cash ETR procedure)
Firm-year observations from CRSP/Compustat merged fundamentals annual from 1999 to 2017	118,260	118,260
Excluding regulated and financial industries (SIC codes between 4900:4999 and 6000:6999)	(37,618)	(37,618)
Excluding Non-U.S. incorporated firms	(13,644)	(13,644)
Excluding observations with negative pre-tax income / ETR's not in 0-1 range and peers not fulfilling those definitions / + calendar year 2017 [5]	(43,395)	(43,983)
Firm-year observations in final data sample (before lag loss)	**23,603**	**23,015**

VIII. Empirical Analysis

8.1 Empirical analysis preface

The empirical analysis is based upon the PLM package in R. As the panel data is unbalanced, basic data management tasks are still conceptually simple, but becomes more cumbersome and error prone. However, PLM handles this perfectly (Croissant & Millo, 2008).

8.2 Summary statistics

[5] Since the ownership data provided by Michael Sinkinson only contains holding information for the first 3 quarters in 2017, we decided to drop the 2017 calendar year from our sample for consistency reasons.

Table 1 represents descriptive statistics for key variables used in our analysis. The sample in panel A consists of a total of 23,603 firm-year observations from 1999 to 2016. The dependent variable is *GAAP_ETR* which is used to measure tax avoidance. At the same time, *GAAP_ETR* has a mean value of 0.349 and a standard deviation of 0.123. *PEER_ETR* has a slightly lower mean value of 0.321 compared to the mean value of *GAAP_ETR*. Nonetheless, *PEER_ETR*'s standard deviation is at a smaller value of 0.039, which means the data clusters around its mean.

For the control variables, *mb* has the mean value of 3.101 with means a firm has on average, a market-to-book ratio of 3.101, with a high standard deviation of 4.331. A low ratio of less than one could indicate that the stock is undervalued, and a higher ratio than one could mean the stock is overvalued. The result implies that most firms' stock in our sample is expensive, with a lower quartile of 1.415 and an upper quartile of 3.656. Total market capitalization (mktcap) has a mean value of 4.102 billion dollars. However, the median value is only 0.821 billion dollars. The significant difference between the mean value and median value can be explained by the big standard deviation (9.980 billion dollars) across firms. Our sample consist of firms of all sizes, including small-cap (i.e. $245 million to $821 million), mid-cap (i.e. $821 million to $2.806 billion), and large-cap ($2.806 billion or more). A firm has on average a ratio of 0.18 in intangible assets to total assets and 0.249 in property, plant, and equipment to total assets. Firms from our sample have an average leverage ratio (*lev*) of 0.171 and R&D (*r_d*) on total assets ratio of 0.049. *HHI* from our example indicates that firms operate in the market without much market concentration.

Please note that the statistics shows values after winsorizing the control variables; *INTAN*, *INST*, *MB, DIFFNOL*, *PPE*, *ROA*, *MKTCAP*, *F_INC*, *EQINC*, *LEV*, and *R_D*, to the 1st and 99th percentiles. Furthermore, all ETR's are filtered to the unit interval between 0 and 1.

We would like to remark that most firms in our sample have a very high concentration of institutional investors in their shareholder register, which on average 66.3%. Firms in the upper quantile have 86.6% of their investors as institutional investors, while institutional investors own around 48,3% of firms in the lower quantile.

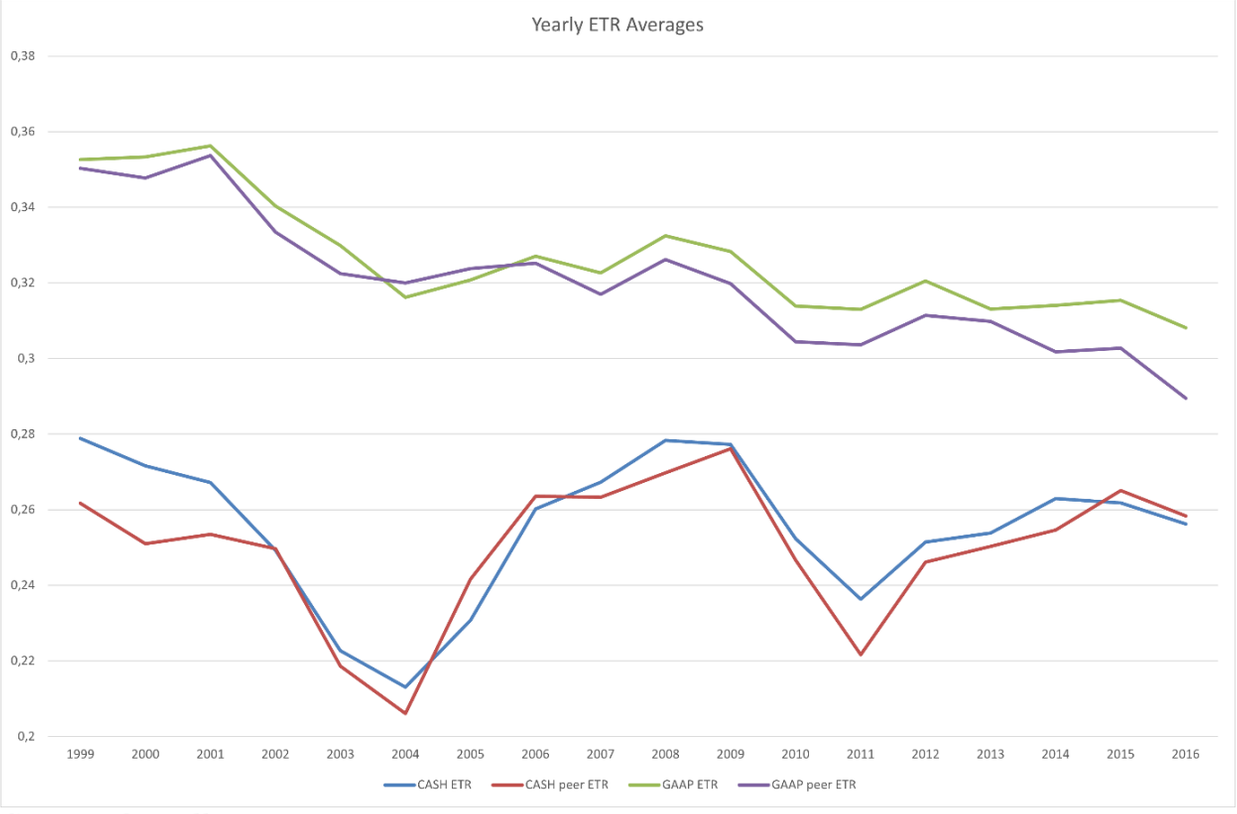
For the *cash ETR* sample in panel B, the results remain similar. However, there are some key differences. Firstly, there is 23,015 firm-year observations due to more missing values of *txpd*

(Income Taxes Paid) compared to *txt* (Income Taxes Total). Secondly the *cash ETR* have a lower mean value and a higher standard deviation. This also contributes to a lower mean and a higher standard deviation of the (cash) peer ETR.

8.3 ETR Graphs

This graph represents the yearly ETR averages across all firms in our net sample. The peer ETR's are calculated as described in the earlier section.

Graph 1: Yearly ETR averages

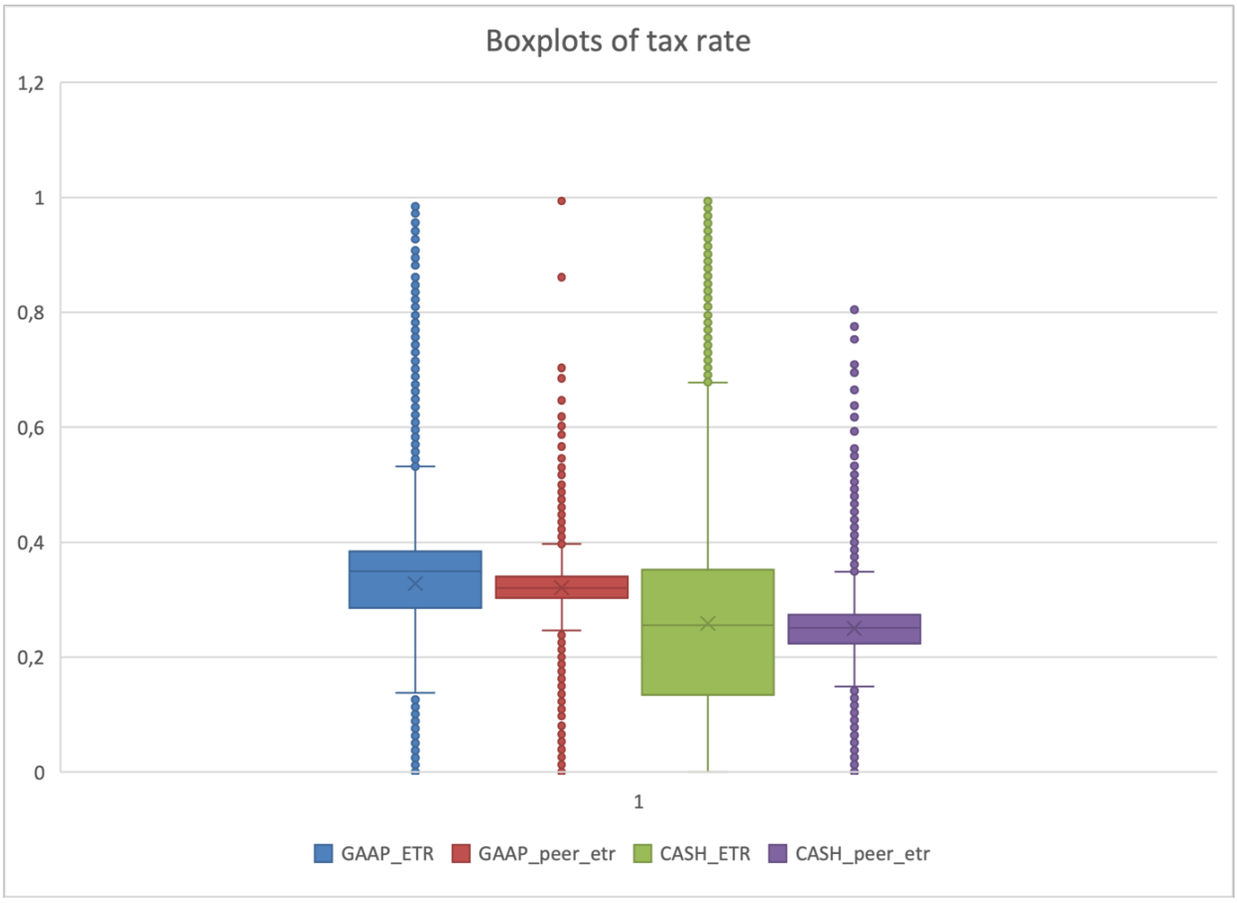


Source: Own illustration

The time-series graph represents the value of the four variables over the observed period from 1999 to 2016. Cash tax rates generally have a lower value than GAAP ETR, and they are also more volatile. We can see there is a declining trend of tax rate over time, with a GAAP tax rate was about 0.35 in 1999 and reduced to around 0.3 in 2016. The cash tax rate of 0.26 reduces from the rate of 0.28 in 1999. The plot demonstrates an exciting result that the yearly peer ETR rates highly correlate with the yearly firm's ETR rates. It gives us high confidence that the peer ETR calculation is free from error, as there is a low correlation between the

dependent and independent variables on firm level, which we detailly examine in the correlation plots below.

Graph 2: Box plot

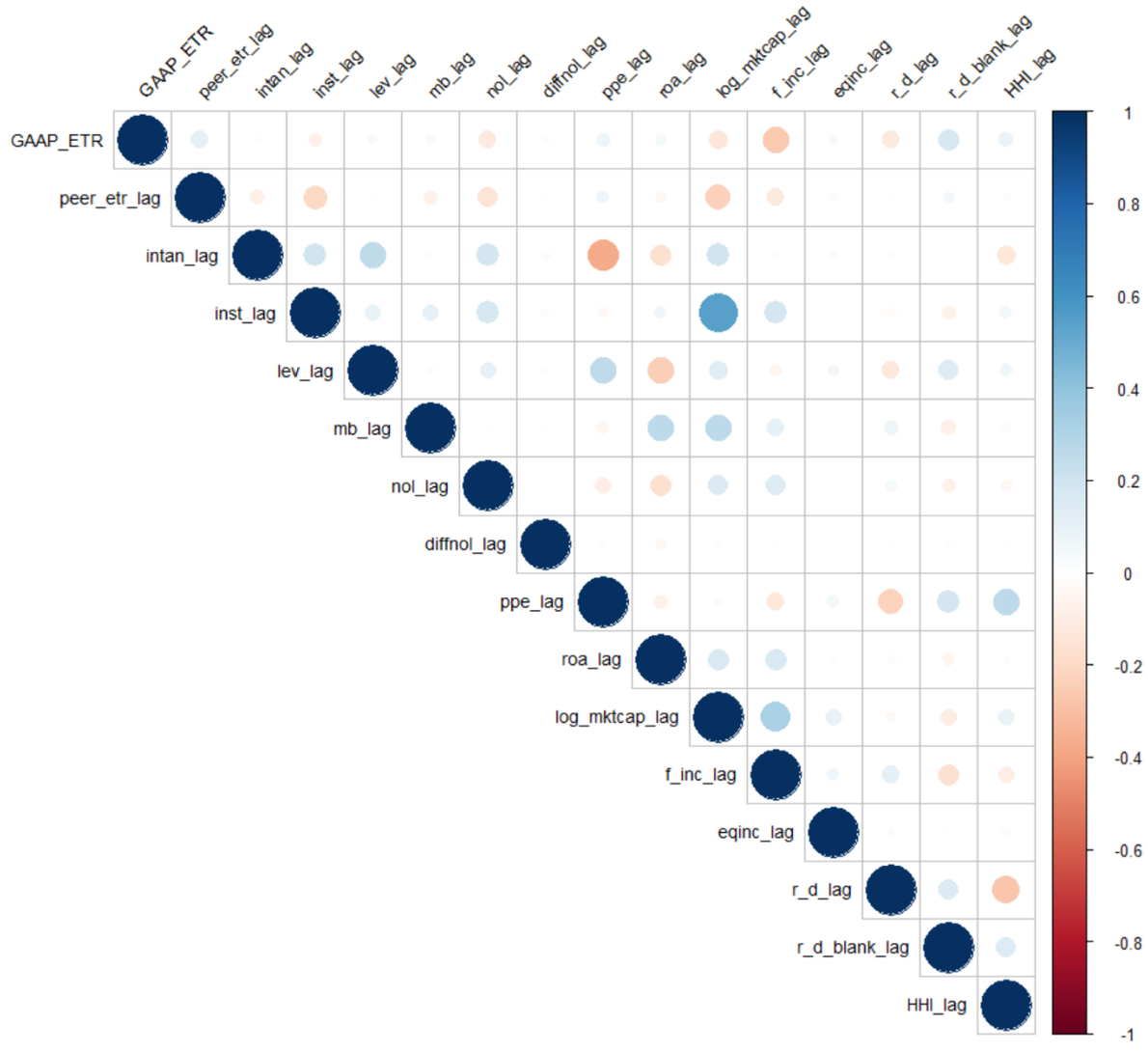


Source: Own illustration

The box plot shows the distribution of our two explainable variables (GAAP ETR and Cash ETR) and two explanatory variables (GAAP Peer ETR and Cash Peer ETR) based on five number summaries: minimum, first quartile Q1, median, third quartile (Q3) and maximum values. GAAP ETR has a median value of around 0.35, and Cash ETR has a median value of approximately 0.25. Peer GAAP ETR has almost identical median value with GAAP ETR, which is a little bit lower than 0.35. Peer Cash ETR's median value is almost identical with Cash ETR's median value. Cash ETR has a wider distribution compared to other variables. Since we have already restricted the values to be between zero and 1, the minimum value showed in the graph is 0 and the maximum value is 1.

8.4 Correlation plots

Plot 1: Correlation matrix for the GAAP ETR sample

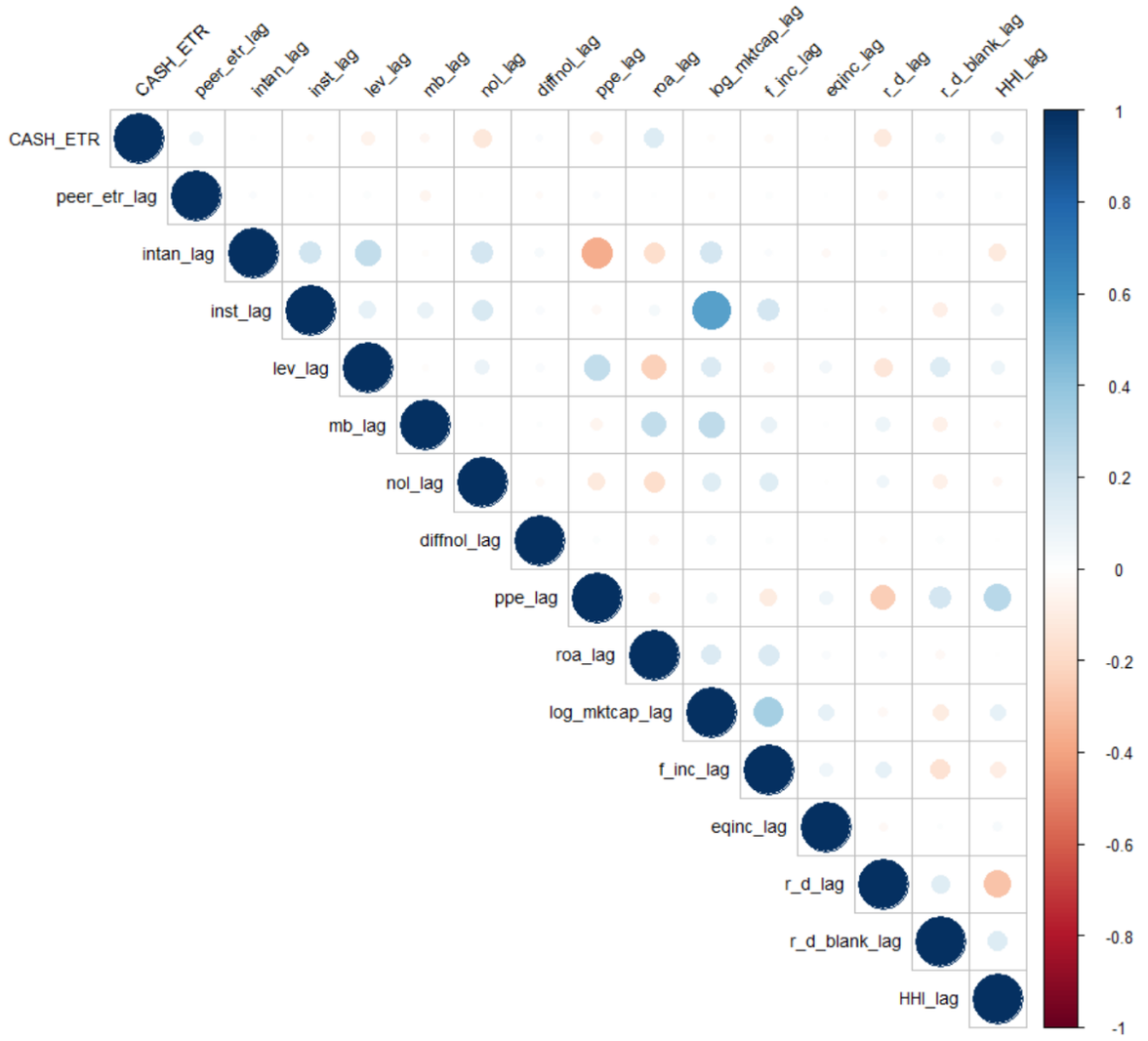


Source: Own illustration

The correlation plot represents correlations in the net sample, using GAAP ETR as the tax avoidance measure. GAAP ETR is in *year t*, and all control variables are lagged one year back. Thus, "_lag" means the variable is in *year t-1*. For variable descriptions, see Appendix A.

The correlations are low overall. However, there are some instances how high correlations, especially between *inst* and *log_mktcap*. Thus, some effect from *log_mktcapt* may bias our *inst* control variable, vice versa. *Inst* is institutional ownership in percentage.

Plot 2: Correlation matrix for the Cash ETR sample



Source: Own illustration

Cash ETR is in *year t*, and all control variables are lagged one year back. Thus, "_lag" means the variable is in *year t-1***.** For variable descriptions, see Appendix A.

Correlations for the cash ETR sample remain similar to the GAAP ETR sample. For specific differences in the number of observations and missing values handling between the two samples, see Appendix C.

8.5 Baseline regression

We estimate the following model to examine how peer firms connected through CIB's affect tax avoidance at focal firms:

$$\begin{aligned} ETR_{i,t} = {} & \beta_0 + \beta_1 PEER\ ETR_{i,t-1} + \beta_2 INTAN_{i,t-1} + \beta_3 INST_{i,t-1} + \beta_4 LEV_{i,t-1} + \beta_5 MB_{i,t-1} \\ & + \beta_6 NOL_{i,t-1} + \beta_7 \Delta\ NOL_{i,t-1} + \beta_8 PPE_{i,t-1} + \beta_9 ROA_{i,t-1} \\ & + \beta_{10} Log\ MKTCAP_{i,t-1} + \beta_{11} F\_INC_{i,t-1} + \beta_{12} EQINC_{i,t-1} + \beta_{13} R\_D_{i,t-1} \\ & + \beta_{14} R\_D\_Blank_{i,t-1} + \beta_{15} HHI_{i,t-1} + INDUSTRY\_FE + YEAR\_FE + e_{i,t-1} \end{aligned}$$

Our baseline regression is given in table 2, panel A and B.

We are left with 17,003 observations when using GAAP ETR as a tax avoidance measure (panel A) and 16,638 when using cash ETR as a tax avoidance measure (panel B). In column 1, there is a simple OLS regression as a multicollinearity assurance. In column 2, we also include year-fixed effects. This removes the impact of variables that vary over time but constant between firms, such as macroeconomic conditions (Hill, Griffiths, & Lim, 2018). The last column (3) includes industry fixed effects using 2-digit SIC. With industry-fixed effects, unobservable heterogeneity that varies across industries is removed.

The most exciting results are in columns (2) and (3). That is, column (2) retains more relevant variation as discussed earlier in the chapter about strategic reactions and time-constant variance at the expense of not including industry fixed effects. In later analysis, we only show tables with both industry fixed effects and yearly fixed effects. However, we do perform robustness tests and event studies with firm fixed effects.

Panel A: GAAP ETR as tax avoidance measure

We see that when peer ETR in year t-1 increases by one percentage point, focal firms follow by increasing their ETR by 0.187 percentage points in year t. The results are significant at the 1% level and similar to the results by Cheng, Sun & Xie (2018), who used other tax avoidance measures.

Furthermore, we do not find significant results regarding institutional ownership in % of common shares outstanding. A higher proportion of institutional ownership in the focal firm

should contribute to the focal firm's tax strategy. In other words, in isolation, a higher share of institutional ownership should have a positive/dampening effect on tax avoidance for focal firms. However, we want to emphasize that the *inst* coefficient was more similar with Cheng, Sun & Xie (2018) before excluding firms with negative pre-tax income, and the sample was larger.

Panel B: Cash ETR as tax avoidance measure

The effect by peer ETR remains significant and is higher than the results using GAAP ETR as a tax avoidance measure. The *peer_etr* coefficient is 0.197 when using industry fixed effects. This finding is weaker than what Cheng, Sun & Xie (2018) found in their robustness analysis in Table 9 Panel D, where they found that the effect was that one percentage point increase in peer ETR in year t-1 gave an increase in focal firm ETR in year t by 0.243 percentage points. However, we use a different approach when calculating Cash ETR since we do not subtract special items from pre-tax income. Furthermore, we differ from Cheng, Sun & Xie (2018) because we exclude observations with negative pre-tax income.

Results from both panels support our H.1.

8.6 Differences in the peer ETR effect: peer firms connected by long versus short-term blockholders

In this analysis, we look at the difference between peer ETR for long-term and short-term blockholders and how they impact focal firms. We want to capture the difference between tax avoidance mimicking towards peer firms based upon how long the investors have been a blockholder in the focal firms.

The calculation is done by considering investors who own at least 5% of current shares outstanding for at least 4 subsequent years. If an investor has been a blockholder for the last 3 years and is also a blockholder in year 4, this blockholder will be classified as long-term in year 4. If the respective blockholder also owns at least 5% of the shares in year 5, the respective blockholder is still classified as long-term in year 5. This is how the count continues until the investor no longer owns 5% of the shares, and when this is the case, the classification will cease. Four new years with 5% holdings are required for this investor to be

reclassified as be reclassified as long-term in the respective firm-year observation. Thus, short-term is the residual classification.

We want to point out that we decided to remove observations of the calendar years 1999, 2000, and 2001. The reason behind this is that we do not know what kind of holdings the investors had in the years before 1999. In other words, 2002 is the first possible year an investor to be classified as a long-term blockholder. The years before 2002 are removed, so we do not wrongly classify short-term investors before 2002 who truly were long-term investors in 1999, 2000 and 2001, respectively.

The reason behind the few firm-year observations is that firms need to have at least one long-term blockholder and one short-term blockholder in the same year. We end up with relatively few observations. In total, we have 5.621 (5,593) observations when using GAAP ETR (CASH ETR) as tax avoidance measure. This means that the power of generalization is somewhat weaker in this regression than our prior regressions. We use a lead-lag model in our analysis, since we are concerned with capturing how focal firm responds to peer firm within a 1-year window.

We implement the following regression model to investigate at the difference between the blockholders classification as short-term and long-term in the focal firms.

$$\begin{aligned} ETR_{i,t} = {} & \beta_0 + \beta_1 PEER\ ETR(long)_{i,t-1} + \beta_2 PEER\ ETR(short)_{i,t-1} + \beta_3 INTAN_{i,t-1} \\ & + \beta_4 INST_{i,t-1} + \beta_5 LEV_{i,t-1} + \beta_6 MB_{i,t-1} + \beta_7 NOL_{i,t-1} + \beta_8 \Delta\ NOL_{i,t-1} \\ & + \beta_9 PPE_{i,t-1} + \beta_{10} ROA_{i,t-1} + \beta_{11} Log\ MKTCAP_{i,t-1} + \beta_{12} F\_INC_{i,t-1} \\ & + \beta_{13} EQINC_{i,t-1} + \beta_{14} R\_D_{i,t-1} + \beta_{15} R\_D\_Blank_{i,t-1} + \beta_{16} HHI_{i,t-1} \\ & + INDUSTRY\_FE + YEAR\_FE + e_{i,t-1} \end{aligned}$$

From table 3 panel A, we see a small difference between long-term peer ETR and short-term peer ETR. The coefficient for long-term peer ETR is 0.085 but is significant only at the 10% level. Furthermore, the coefficient for short-term peer ETR is 0.076, and significant at the 5% level. Hence, when using GAAP ETR as tax avoidance measure, our H.2.1 is partially supported because of significance levels.

From panel B, we see a significant effect in terms of short-term peer ETR when using Cash ETR as measure of tax avoidance. The coefficient of *peer_etr_long* is not significant. Furthermore, *peer_etr_short* coefficient is 0.186 and significant at the 1% level. Focal firms follow the portfolio of short-term blockholders to a much greater extent than the long-term blockholders portfolio when it comes to Cash ETR. Thus, H.2.1 is supported as short-term blockholders should have an more impact on focal firm's tax adjustment.

There is still one caveat with these results in mind. Chen, Harford & Li (2007) argue that long-term institutions focus on monitoring and influencing rather than trading for profit, while investors with a short-term horizon are better at predicting short-term profits (Yan & Zhang, 2009). Furthermore, our results indicate that firms do not try to deviate from their long-term blockholders other portfolio firms regarding cash ETR. This gives us some concern that the peer firms connected via short-term blockholders at the focal firms, which drives the effect, might be due to similar firm-level characteristics between the firms in the peer group due to "ex-ante" preferences. Indicating that short-term blockholders are better at predicting levels of tax avoidance when they build their portfolios.

Yet, we argue there is a paradoxical mechanism if the investors build their portfolios based on "ex ante" preferences for tax avoidance. This mechanism speaks against the endogeneity problem. To illustrate this point, consider the following illustration:

"*Investor A buys into a firm and leaves 3 years later, while investor B buys into another firm and remains in that firm. Investor A represents a short-term investor, while investor B represents a long-term investor. Both build their portfolio based on "ex-ante" preferences regarding tax avoidance. If investor A is satisfied with the company's tax avoidance policies according to its "ex-ante" preferences, there is less reasons to leave the firm for other investment opportunities, everything else being equal. Similarly, there is greater reason for investor B to relocate his holdings if the investor is very dissatisfied with the firm's tax avoidance policies in accordance with his "ex-ante" preferences, all other things being equal.*"

Thus, if investors build their portfolio based on their "ex-ante" preferences regarding tax avoidance, there is less reason to believe that *Peer_etr_portfolio_short* (and a greater reason to believe that *Peer_etr_portfolio_long)* represents such preferences.

However, we look further into the endogeneity problem in our event studies below.

8.7 Differences in the peer ETR effect: peer firms with long versus short-term blockholders

We want to examine whether there is any difference between the blockholders portfolio in terms of short-term and long-term holdings. To investigate this further, we use the same method to separate blockholders into short-term and long-term. Unlike other previous analysis, we look at two different groups of portfolios for the same individual blockholders. We spilt their portfolio into different groups. We have a sub-portfolio where the blockholders short-term holdings are included, while the other sub-portfolio contains the blockholders long-term holdings.

Thus, we want to see whether the blockholders long/short-term classification in peer firms matter regarding tax avoidance mimicking.

A long-term blockholder is still defined as an institutional investor which have been a blockholder in the same firm for at least in 4 subsequent years, and short-term is defined as the opposite. We exclude the data from years 1999, 2000 and 2001 as discussed earlier. The analysis aims to look at the effect of different portfolio groups belonging to same specific blockholders, regardless of the blockholders classification in the focal firms.

We implement the following regression model to investigate at the difference between the blockholders classification as short-term and long-term in the peer firms.

$$
\begin{aligned}
ETR_{i,t} = \ & \beta_0 + \beta_1 PEER\ ETR(long\_term\ portfolio)_{i,t-1} \\
& + \beta_2 PEER\ ETR(short\_term\ portfolio)_{i,t-1} + \beta_3 INTAN_{i,t-1} + \beta_4 INST_{i,t-1} \\
& + \beta_5 LEV_{i,t-1} + \beta_6 MB_{i,t-1} + \beta_7 NOL_{i,t-1} + \beta_8 \Delta\ NOL_{i,t-1} + \beta_9 PPE_{i,t-1} \\
& + \beta_{10} ROA_{i,t-1} + \beta_{11} Log\ MKTCAP_{i,t-1} + \beta_{12} F\_INC_{i,t-1} + \beta_{13} EQINC_{i,t-1} \\
& + \beta_{14} R\_D_{i,t-1} + \beta_{15} R\_D\_Blank_{i,t-1} + \beta_{16} HHI_{i,t-1} + INDUSTRY\_FE \\
& + YEAR\_FE + e_{i,t-1}
\end{aligned}
$$

From table 4 panel A, we see that the focal firms' follows their blockholders short-term sub-portfolio, to a greater extent than the long-term sub-portfolio. An increase (decrease) in peer ETR for the short-term portfolio by 1% point, leads to focal firm increase (decrease) its GAAP ETR by 0.108 % point the following year. While an increase (decrease) in peer ETR for the long-term portfolio by 1% point, leads to focal firm increase (decrease) its GAAP ETR by 0.065 % point the following year. Both results are significant at the 1% level.

In panel B, we repeat the analysis with Cash ETR. The ETR coefficient for short-term portfolio firms is 0.193 and significant at the 1% level, while the ETR coefficient for long-term portfolio firms is 0.066 and significant at the 5% level.

Results from both tax avoidance measures supports our H.2.2. The tax avoidance effects are stronger in the short-run then the long-run.

IX. Supplementary empirical analysis

9.1 Differences in the peer ETR effect: peer firms connected by the largest blockholder versus peer firms connected by other blockholders

A problem with baseline analysis is that a focal firm can have many different CIBs each year. In this way, we cannot know whether the different CIBs for each company have the same preferences for tax planning. Cheng, Sun & Xie (2018) refer to this problem as «the concern of multiple common institutional blockholders.» It is essentially the same issue we looked at when analyzing differences between long-term and short-term blockholders.

Cheng, Sun & Xie (2018) argue in their study that the largest common blockholders have the greatest ability to influence a firm's tax planning. They restrict their sample by only looking at the largest CIB's.

In our analysis, we include both the largest CIB and the remaining CIBs in the focal firms. Hence, we look at the difference between how these two different groups of peer firm ETR's might explain changes in focal firms ETR. In this way, we can see the different impacts between the two groups. Peer ETR calculated based on peer firms connected through the

largest blockholder in the focal firms is called *peer_etr_largest*, while for the remaining blockholders peer firms are included in the variable *peer_etr_rest*.

We calculate peer ETR in two rounds. First, we subset the database by looking only at the CIBs in the focal firm, which is the largest in the respective year. Then the weighted market-capitalized peer ETR is calculated for the entire portfolio of the respective CIBs in the given year, excluding the focal firm itself. The procedure is done for all the different firms per year, across all years in the sample.

Next, we follow the same procedure, but we only look at blockholders which are not the largest in focal firms. We end up with two different datasets with different peer ETRs. Firstly, this gives us a table with peer ETR for the smallest CIBs in a focal firm, and secondly, a table with peer ETR for the largest CIBs for the focal firm.

Finally, the datasets are merged to only firm-year observations that contain both peer ETR for the largest CIBs and peer ETR for the smallest CIBs. In other words, we only look at the observations where the focal firm both has a CIB, which is the largest CIB, but where the focal firm also has some remaining CIBs. For example, an observation with a calculated value of *peer_etr_largest* and a missing value of *peer_etr_rest* will be dropped from the analysis. In this way, we capture the difference between the two groups.
We examine the difference between peer firms connected through the two groups of blockholders at the focal firms. Hence, we examine which portfolio impacts focal firms ETR with the following regression:

$$
\begin{aligned}
ETR_{i,t} = \ & \beta_0 + \beta_1 PEER\ ETR(largest)_{i,t-1} + \beta_2 PEER\ ETR(rest)_{i,t-1} + \beta_3 INTAN_{i,t-1} \\
& + \beta_4 INST_{i,t-1} + \beta_5 LEV_{i,t-1} + \beta_6 MB_{i,t-1} + \beta_7 NOL_{i,t-1} + \beta_8 \Delta\ NOL_{i,t-1} \\
& + \beta_9 PPE_{i,t-1} + \beta_{10} ROA_{i,t-1} + \beta_{11} Log\ MKTCAP_{i,t-1} + \beta_{12} F\_INC_{i,t-1} \\
& + \beta_{13} EQINC_{i,t-1} + \beta_{14} R\_D_{i,t-1} + \beta_{15} R\_D\_Blank_{i,t-1} + \beta_{16} HHI_{i,t-1} \\
& + INDUSTRY\_FE + YEAR\_FE + e_{i,t-1}
\end{aligned}
$$

Our findings in Table 5 remove some of the problems that different CIBs may have different tax planning preferences, and we find that the effect for *peer_etr_largest* is still positive and significant. In other words, the effect goes in the same direction as *peer_etr_rest*. The findings

do not fully support the argument by Cheng, Sun & Xie (2018) that "the largest common institutional blockholder is likely to influence firms' tax avoidance policies to a greater extent than other CIBs." Thus, we like to empathize that *peer_etr_rest* still contains a large group of peer firms.

From panel A we see that the coefficient of *peer_etr_largest* is 0.076 while *peer_etr_rest* is 0.163 when using GAAP ETR as a tax avoidance measure. These coefficients are significant at the 1% level.

In panel B, the difference between the coefficients is still similar in magnitude. However, with Cash ETR as a tax avoidance measure, the ETR adjustment towards *peer_etr_largest* is greater in magnitude. Both groups of peer ETR are significant at the 1% level.

This means that the largest CIBs do indeed have a great impact on a firm's tax avoidance.

X. Empirical analysis: robustness tests

10.1 Firm fixed effects

We test whether the results are robust to firm fixed effects. The main reason not to use this approach is that it removes time-constant variation, which we want to capture for this research. Thus, some relevant variations are removed. We refer to the discussion about strategic reaction for a better explanation.

Generally, the main reason for using fixed effects at fixed levels is that unobservable heterogeneity can lead to bias in the estimators. A correlation between an unobservable variable and the variable we are interested in will lead to bias in the variable we are interested in (i.e., bias in peer ETR). With fixed effects at the individual level, we remove the sources of bias from unobservable time-constant variables (Collischon & Eberl, 2020).

We perform the following regression:

$$ETR_{i,t} = \beta_0 + \beta_1 PEER\ ETR_{i,t-1} + \beta_2 INTAN_{i,t-1} + \beta_3 INST_{i,t-1} + \beta_4 LEV_{i,t-1} + \beta_5 MB_{i,t-1} + \beta_6 NOL_{i,t-1} + \beta_7 \Delta\ NOL_{i,t-1} + \beta_8 PPE_{i,t-1} + \beta_9 ROA_{i,t-1} + \beta_{10} Log\ MKTCAP_{i,t-1} + \beta_{11} F\_INC_{i,t-1} + \beta_{12} EQINC_{i,t-1} + \beta_{13} R\_D_{i,t-1} + \beta_{14} R\_D\_Blank_{i,t-1} + \beta_{15} HHI_{i,t-1} + FIRM\_FE + YEAR\_FE + e_{i,t-1}$$

The results can be found in table 6, panel A and B. With firm fixed effects, we only look at the variation within the firms that varies over time but not at the variation that only varies across the companies and is constant. The regressions also include yearly fixed effects, and we remove variation over time, which is constant between the firms. In this way, unobservable variables that only vary over time (within variance) or vary only across companies (between variation) are omitted from the regression (Hill, Griffiths, & Lim, 2018). The effect of peer ETR on the focal firm's ETR remains significant and high in magnitude, both when using GAAP ETR and cash ETR as tax avoidance measures. See table 7, panels A and B.

10.2 Peer firm restriction based on SIC

There are several different methods of defining peer groups of a focal firm. Previous studies in tax avoidance and peer effects define peer groups as firms within the same industry or different variants of this definition (Bird & Karolyi, 2017; Bird & Karolyi, 2018).

Our approach, similar to Cheng, Sun & Xie (2018), identities peer companies by looking at each firm's CIBs in the calendar year and then weighing ETRs based on a common investor base, excluding the focal firm itself. Our baseline analysis does not know whether the focal firm and its peer firms operate in the same industry.

In the study by Heitzman & Ogneva (2019), tax policy is a potential specific channel for the systematic risk of tax policies will change, affecting after-tax cashflows. If investors believe that industry-driven tax planning increases the firm's non-diversifiable cash flow shocks, they require a higher risk premium for firms that belong to a high tax planning industry. Then it is natural that some of the effects come from the fact that focal firm and peer firm are within the same industry and have the same exogenous characteristics since an investor may have investments in several companies within the same industry. Firms organize their tax planning

strategies around industry-specific incentives and opportunities, and they also respond to other firms within the same industry because of competitive and strategic reasons (Armstrong, Glaeser, & Kepler, 2019; Bird, Edwards, & Ruchti, Taxes and Peer Effects, 2018; Kubick, Lynch, Mayberry, & Omer, 2015). Tax laws are industry-centric; hence, tax enforcement structures around taxpayers' operating industries (Heitzman & Ogneva, 2019)

As described in the section about the reflection problem, it may be that a firm follows the dominant tax strategy within the industry. In this way, we lose track of whether the effect is driven by investors or because focal firms and peer firms are in the same industry. An overview of peer firms operating in the same industry is presented in table 1.2.

We examine whether the investors are driving the effect, but not the companies being in the same industry. Our baseline analysis contains components of both effects. We want to see if the effect still holds if we force peer firms that share the same industry with a focal firm out of the CIBs portfolio. We are aware that we lose several ETR contributions from investors with holdings in both focal and peer firms, where peer firms have the same industry code as the focal firm. The effect will contain both a component of a direct effect from the investors and an indirect effect of focal firms and peer firms being in the same industry. This study contributes to minimizing the later effect.

Because of this reason, we consider this analysis to be a robustness test to our baseline regression. We expect the effect to fall somewhat as the industry codes go from 4-digit to 2-digit. The 1-digit SIC represents a division, while the 2-digit SIC codes represent a significant group, the 3-digit stands for the industry group, and the 4-digit present industry sector. The effect of peer firms on the focal firm decreases as we use 4-, 3- and 2-digit SIC codes, respectively.

When weighting of ETR according to the blockholders market capital portfolio, we exclude companies that share the same SIC-code with the focal firm from the weighting. In other words, total market capitalization in the blockholders' portfolio will not include the excluded firms so that the weighting is not skewed. Each firm's market capitalization is aggregated to 100% and then weighted as its share of the total portfolio.

We estimate the following three models, where *peer_ETR* is calculated as stated below. For the first model, peer firms with the same 4-digit SIC as the focal firm are excluded. For the second model, peer firms with the same 3-digit SIC are excluded. Lastly, peer firms with the same 2-digit SIC are excluded:

$$\begin{aligned} ETR_{i,t} = {} & \beta_0 + \beta_1 PEER\ ETR_{i,t-1} + \beta_2 INTAN_{i,t-1} + \beta_3 INST_{i,t-1} + \beta_4 LEV_{i,t-1} + \beta_5 MB_{i,t-1} \\ & + \beta_6 NOL_{i,t-1} + \beta_7 \Delta\ NOL_{i,t-1} + \beta_8 PPE_{i,t-1} + \beta_9 ROA_{i,t-1} \\ & + \beta_{10} Log\ MKTCAP_{i,t-1} + \beta_{11} F\_INC_{i,t-1} + \beta_{12} EQINC_{i,t-1} + \beta_{13} R\_D_{i,t-1} \\ & + \beta_{14} R\_D\_Blank_{i,t-1} + \beta_{15} HHI_{i,t-1} + INDUSTRY\_FE + YEAR\_FE + e_{i,t-1} \end{aligned}$$

From table 7, the peer effect on the focal firm's decreases as the SIC-code restriction becomes coarser and coarser. Panel A (*GAAP_ETR*) shows a 1%-point increase in *peer_ETR* resulting in the focal firm's *GAAP_ETR* increase with 0.164% point in the following year, using a 4-digit SIC restriction. In the 3-digit SIC restriction and 2-digit SIC restriction analysis, the corresponding coefficients are 0.158 and 0.165, respectively. The results remain significant at the 1% level.

In panel B, the coefficients for 4, 3, and 2-digit SIC restrictions are 0.187, 0.178, and 0.174, respectively. All the results remain significant at the 1% level.

In conclusion, our finding indicates that the focal firm's tax avoidance is not only driven by the indirect component since peer firms operate in the same industry as the focal firm. Hence, this result supports that the managers are more likely to drive the effect. Thus, we can conclude that some of the firms in peer groups being in the same industry do not interfere with our main research question: *"Do firms engage in similar levels of tax avoidance if the same blockholders own them?"*

10.3 Event study on shocks in peer ETR's

We study the impact of peer firms' tax rate shocks measured by the reaction of the focal firm indices of several tax shocks observed in 1999-2016. We apply the event study methodology to the firm-specific financial information and other factors that measure tax avoidance

behaviour as control variables. The approach will allow us to identify the reactions of the focal firm's tax rate given exogenous events, i.e., tax shocks in the peer firms.
Cheng, Sun & Xie (2018) points out the possibility that investors might build their portfolios based on "ex-ante" preferences for tax avoidance. *Peer_etr* is endogenous, and the firms do not follow each other based upon CIB mechanisms. In comparison, the two CIB mechanisms are the possibility that the investors influence the firm's management or the second possibility that management looks towards their investor's other portfolio firms.

This regression aims to solve the endogeneity problem, using exogenous ETR shocks in peer firms. We constructed two arguably exogenous variables, *Negative shock year* (The year of the negative shock peer ETR) and *Positive shock year* (the year of positive shock in peer ETR), through several parsimonious steps. A change in peer tax rate is measured by taking peer ETR in the firm-year (*t*) minus peer ETR in the year before (*t-1*).

We set up a dummy variable equal to 1 if the peer ETR is either larger or equal to 5% and smaller than 10% in absolute terms. Then we do the same procedure with a change in peer ETR larger than 10% in absolute terms. The dummies that capture shocks in year t are lagged and led to years before and after the shock. We also exclude the shock year (*t+1*). Hence, the dummy for the year before the shock year is in the reference group. Furthermore, we bin the remaining years in the dummies "*5+ years after (before) neg(pos) shock"*. Our set of control variables are included, and we make use of firm fixed effects and year fixed effects to get the most unbiased results.

We conduct the following regression:

$$\begin{aligned} ETR_{i,t} = {} & \beta_0 + \beta_1\, YEAR\_DUMMIES_{i,t} + \beta_2 INTAN_{i,t} + \beta_3 INST_{i,t} + \beta_4 LEV_{i,t} + \beta_5 MB_{i,t} \\ & + \beta_6 NOL_{i,t} + \beta_7 \Delta\, NOL_{i,t} + \beta_8 PPE_{i,t} + \beta_9 ROA_{i,t} + \beta_{10} Log\ MKTCAP_{i,t} \\ & + \beta_{11} F\_INC_{i,t} + \beta_{12} EQINC_{i,t} + \beta_{13} R\_D_{i,t} + \beta_{14} R\_D\_BLANK_{i,t} + \beta_{15} HHI_{i,t} \\ & + FIRM\_FE + YEAR\_FE + e_{i,t} \end{aligned}$$

Both panel A and panel B of Table 9 provides exciting results when using GAAP ETR and cash ETR as measurements of tax avoidance. For *GAAP_ETR* (panel A), there is a clear pre-trend before a negative shock in peer ETR between 5% and 10%. In the years before the

shock, there is an abnormally high GAAP ETR in the focal firm. In the event year, the focal firm's ETR already decreases. The decrease in GAAP ETR in one firm in the peer group also contributes to a shock in ETR across the peer group.

Let us consider three firms, firm *A*, *B,* and *C*, as a peer group. A decrease in GAAP ETR for firm *A* will contribute to a shock in the peer group of firm *B* (containing firm *A* and *C*). Furthermore, the decrease in GAAP ETR for firm *A* will contribute to a shock in the peer group of firm *C* (containing firm *A* and *B*). Thus, there is always a rotation in the peer groups.

Firms do indeed follow an arguably exogenous negative shock in peer ETR for GAAP ETR. The results persist over the years, assuring us that the peer ETR is exogenous and not endogenous because investors might build their portfolios based on "ex-ante" preferences for tax avoidance.

The same logic goes for positive shock in peer ETR when using GAAP ETR as a tax avoidance measure. However, the results do not persist in the same magnitude as the negative peer ETR shocks. When the shock in GAAP peer ETR is above 10% in absolute terms, we do not find similar results. However, there are also more minor tax shocks in this range.

We conduct the same analysis using cash ETR as a tax avoidance measure in panel B of Table 8. When the positive shocks in cash peer ETR are between 5-10%, we see a clear pre-trend and post-trend. Before a positive shock in cash peer ETR, firms have an abnormally low cash ETR. In the shock year, they have an abnormally high cash ETR, which contributes to the shock in the peer groups. In the years following the positive shocks, there is a clear post-trend with abnormally high cash ETR. The same story goes when there are shocks in cash peer ETR greater than 10%. We do not see a pre-trend, but the post-trend seems systematic. Thus, firms do indeed follow positive cash ETR shock in the peer group.

As for negative cash peer ETR shock, we have some exciting results. When the negative shocks are between 5-10% in absolute terms, we do indeed see a pre-trend with abnormally high cash ETR. There is also an indication of this when the shock is more significant than 10% in absolute terms. The abnormally high cash ETR is not the case in the shock year, contributing to the negative shocks themselves.

However, in the years after there are some interesting effects. We can see abnormally high cash ETR in the years following negative cash ETR shocks in peer groups. This brings us to the idea that investors might change their investment when there is a negative cash ETR shock in their portfolio firms. In the next section, we analyse what investors do with their portfolio weights using a similar event study approach.

Below, we discuss some findings in the literature to explain what might be going on.

As discussed above, some literature explains why we see an increase in the firms' ETR after the peer firms are exposed to a negative shock. Dobrzynski (1993) and Monks & Minow (2011) argue that large shareholders monitor and discipline managers to ensure that their investment strategies are consistent with maximizing long-term value instead of short-term earning goals. In our context, we can assess developments in tax avoidance and interpret the results based on what provides the highest firm value in the long run. Chen, Chen, Cheng & Shevlin (2010) argues that long-term institutional shareholders must consider a trade-off between tax savings through tax avoidance strategies and associated reputation damage and penalties if later ruled improper.

Thus, one mechanism could be that if firms included in an investor's portfolio experience a negative cash ETR shock, the investor will react by reducing their investment in the respective firms. Hence, the investors are trying to normalize tax avoidance within their portfolios to a more normalized level. In this way, abnormal changes to tax avoidance do not reflect poorly on the investor due to reputation damages. In conclusion, we argue that the effects are due to long-term institutional shareholders concerned about reputation damage and try to reverse abnormal levels of tax avoidance on a portfolio level.

For descriptive statistics over the shocks in our sample, see appendix B.

10.4 Event Study on shocks in ETR's on the investor-level

Cheng, Sun & Xie (2018) sheds light on the endogeneity problem in which the blockholders chooses to hold firms with similar tax avoidance. Thus, the blockholders build their portfolios based on "ex ante" preferences.

Cheng, Sun & Xie (2018) describes this potential mechanism as the "self-selection hypothesis". In the previous event study, we introduced exogenous shocks in the peer ETR's to solve the endogeneity problem. However, it is also worth investigating how blockholders build their portfolios on a "investor level".

To further investigate whether institutional blockholders may build their portfolio based on "ex-ante" preferences for tax avoidance, we conduct one final analysis. In this analysis, we study whether shocks in the focal firm's ETRs contribute to how blockholders weigh their holdings in those firms.

This event study is similar to the one in the previous subsection. However, there are a few differences to consider. Firstly, we are on the investor level. Thus, every firm-investor combination are the cross-sectional units, whilst the time unit remains the same. Furthermore, we do not look at shocks in peer firm's ETR, but the ETR's of the individual firms themselves.

Because, unlike the peer ETR, focal firm's ETR is not weighted average of multiple ETR's. Thus, focal firm's ETR in nature will be more volatile. Therefore, we increase the shock-range to 5%-15% changes in ETR for column (1) in table 10. For column (2) the shocks are greater than 15% in absolute terms.

We refer to table 3 in chapter 4 about the ETR weighting process for the explanation for this analysis. We examine whether shocks in ETR's (column 5 of Figure 2.2) contribute to changes in the weights (column 8 of Figure 2.2). Simply put, in the event study, we want to capture if the blockholders sell/buys shares as the ETRs of those firms change.

As for reference. Figure 2.2:

CIK (1)	CUSIP (2)	Shares (3)	PRCC (4)	ETR (5)	MKTCAP (6)	Total MKTCAP (7)	(6)/(7) = (8)	Weighted Peer_ETR (9)
1	B	80	7	0.33	560	1955	0.2864	0.094526854
1	C	90	6	0.34	540	1955	0.2762	0.093913043
1	D	95	9	0.35	855	1955	0.4373	0.153060905
						Sum Peer_ETR		0.341508951

We conduct the following regression analysis on the investor level:

$$ETR_{i,t} = \beta_0 + \beta_1 YEAR\_DUMMIES_{i,t} + \beta_2 INTAN_{i,t} + \beta_3 INST_{i,t} + \beta_4 LEV_{i,t} + \beta_5 MB_{i,t} + \beta_6 NOL_{i,t} + \beta_7 \Delta NOL_{i,t} + \beta_8 PPE_{i,t} + \beta_9 ROA_{i,t} + \beta_{10} Log\ MKTCAP_{i,t} + \beta_{11} F\_INC_{i,t} + \beta_{12} EQINC_{i,t} + \beta_{13} R\_D_{i,t} + \beta_{14} R\_D\_BLANK_{i,t} + \beta_{15} HHI_{i,t} + INDIVIDUAL\_FE + YEAR\_FE + e_{i,t}$$

When using GAAP ETR as the tax avoidance measure, we experience 4,100 negative ETR shocks and 3,491 positive ETR shocks across the sample in the 5%-15% shock range. In the 15%+ shock range, we experience 2,798 negative ETR shocks and 2,997 positive ETR shocks.

When using cash ETR, we experience 6,686 negative ETR shocks and 7,573 positive ETR shocks in the 5%-15% shock range. Whereas the 15%+ shock range gives 4,568 negative ETR shocks and 5,844 positive ETR shocks.

Table 9 panel A shows a portfolio weight decrease the year after a negative shock in firms GAAP ETR, but it does not seem to be systematic as it reverses the following year.
Panel B is more interesting. When there is a negative shock in firms' cash ETR between 5-15%, the blockholders reduce their weights. This effect persists and seems systematic. Furthermore, this might explain why we are seeing abnormally high cash ETR following negative peer ETR shocks in the previous subsection. That is, the blockholders might change their portfolios towards firms with higher levels of cash ETR. Also, the reaction could also serve as a signal to management to reduce their aggressive tax avoidance, encouraging management to engage in less aggressive tax avoidance.

In conclusion, the results also shed light on the "self-selection hypothesis". However, in sum the results do not seem to be great enough in magnitude to conclude that investors build their portfolios based on "ex-ante" preferences for tax avoidance [6]. Furthermore, our results indicate that institutional blockholders seem to dislike aggressive tax avoidance levels. This might be an interesting topic for further research.

[6] In unreported regression, we did not find a significant effect by ETR's on the portfolio weights with individual fixed effects, using weights as dependent variable and ETR's as independent variable.

XI. Limitations and Alternatives

As with many studies, the design of the current study is subject to limitations.

Although GAAP ETR is widely used in tax studies, it has posted some major drawbacks. First of all, it is based on only annual data, therefore causes significant year-to-year variation in the annual effective tax rate and undefined tax rates due to negative denominators that can obscure inferences about tax avoidance (Bird, Edwards, & Ruchti, 2018). We observe fully the firm's tax avoidance behaviour allows endogenous variables to capture all firm's accelerating deductions and deferring income for tax purposes relative to book purposes, reducing current taxes but increasing deferred taxes.

In this case, GAAP ETR includes both current and deferred taxes; it will not fully reflect such forms for tax avoidance. A common method used by multinational enterprises is offering stock options to their employees. Firms get deductions when stock options are exercised. Still, GAAP ETR recognized no expense at either grant date or exercise date, which creates a permanent difference between book and tax accounting. Tax benefits are then added directly to equity rather than reducing current tax expenses. Current tax expenses are overstated compared to actual tax paid for firms which such a kind of tax deductions. Another problem can is, by using GAAP ETR, book accrual within tax expense such as valuation allowance and tax contingency reserve has not been considered in our model.

It is a simplification of how we calculate the portion of current shares outstanding owned by investors. We take the number of shares from *common ownership data* (by Michael Sinkinson) registered per quarter and investor and use this to calculate the annual average of shares. The market capitalizations are the number of shares is multiplied with *price per share close (fiscal)*. At the same time, the shares at the end of the accounting period times *price per share close (fiscal)*, representing the actual market capitalization.

Furthermore, we have experienced several instances that the sum of quarterly shares for all investors in a company from *common ownership data* has exceeded the *cshoq* (current shares outstanding quarterly) registered in *Compustat/CRSP-merged quartal fundamentals* for the same firm. This is also the case when summing quarterly shares data from *common ownership data* to yearly values and compare those values against *csho* (current shares outstanding

annual) from the *Compustat/CRSP-merged annual fundamentals* database. This may lead to measurement errors in the weighting of peer ETR and the calculation of institutional ownership in percent.

We also have limited peer firms to firms that share a common institutional blockholder (Cheng, Sun, & Xie, 2018). Thus, a blockholder is defined as an investor who owns at least 5% of current shares outstanding for at least one quarter throughout the year. It could also be interesting to look at peer firms connected by institutional owners who own less than 5% of current shares outstanding in both firms.

Since we use effective tax rates as tax avoidance measures, we follow other researchers and delete loss year observations, where pre-tax income is negative or equal to zero. Therefore, we are only using a subsample of profitable firms. This is a limitation, and we are indeed exposed to «truncation bias» (Henry & Sansing, 2018). We suggest further researchers should consider using other measures of tax avoidance and not ETR's.

Furthermore, studies like Armstrong, Glaeser & Kepler (2019) winsorize ETRs to the unit interval between 0 and 1, while we have discretionarily chosen to drop firms with ETR's outside the unit interval, both before and after calculation of peer ETR. It means that we have fewer measurement errors in our ETR's but at the expense of a more unbalanced panel data model. Finally, we will emphasize that calendar years and financial years do not overlap perfectly, and that the fiscal year is compiled based on which calendar year most of the fiscal months are within.

XII. Conclusion

This paper is intended as a supplement to Cheng, Sun & Xie (2018) and we follow many of the same methods. First, like Cheng, Sun & Xie (2018), we look at whether firms with common institutional ownership engage in similar levels of tax avoidance. Furthermore, we follow their methodology using a lead-lag model to control events that affect both focal firms and peer firms within the same year.

Unlike Cheng, Sun & Xie (2018), we perform several alternative regressions. We differentiate between investors who have owned 5% of current shares outstanding in firms for at least four

subsequent years and investors who do not. We have indications that tax avoidance mimicking is stronger towards peer firms where the investors recently placed their investments. In addition, we look at whether focal firms mimic ETR's of peer firms connected by their most significant investor. The result is partly supportive of Cheng, Sun & Xie (2018), who claims that the largest CIB's are likely to influence a firm's tax avoidance more than other CIB's. We agree with the statement, but whether the peer effect is more significant between two groups of peer firms has a lot to do with the number of peer firms connected inside the two different groups. We see that the peer effect of the largest CIB's is significant and goes in the same direction as the peer effect of other CIB's. Therefore, we can conclude that the largest CIB's at least significantly impact the company's tax avoidance.

Furthermore, the results are robust for firm fixed effects, and we know that latent firm-level characteristics do not merely drive the peer effect. Another problem, in contrast to Cheng, Sun & Xie (2018), is that we have more firms in peer groups operating in the same industry. To see if the results are not driven by the fact firms operate in the same industry, we have developed simple techniques to see if such effects do not drive the results. We did this by omitting peer firms with the same SIC as the focal firm in the peer ETR calculation. We find that peer ETR is still significant with such peer firms excluded from the peer ETR calculation.

To investigate the endogeneity problem further, we see how firms react to arguably exogenous shocks in peer ETR. When we use GAAP ETR as a tax avoidance measure, we indicate that peer ETR is exogenous since companies respond to shocks in peer ETR. It applies to both negative and positive peer ETR shocks.

When we use cash ETR as a tax avoidance measure, the evidence is not as clear. The result is because we see an abnormally high cash ETR in the year following negative peer ETR shocks.

Finally, we conduct an event study on the investor level to examine whether blockholders adjust their portfolio weights when there is a shock in the focal firm's ETR. When using GAAP ETR as a measure, we see no reaction. This result gives us further support that peer ETR based on GAAP ETR measure is exogenous. However, when using cash ETR as the measure for tax avoidance, we see that blockholders adjust their portfolios when there is a negative cash ETR shock for their portfolio firms. It helps us interpret why we experience

positive reactions to negative (cash) peer ETR shocks. Because when investors adjust their weights, possibly buying into firms with less aggressive tax avoidance, the peer ETR will automatically increase. The investor reaction could also signal to management to reduce their aggressive tax avoidance, making cash ETR increase further after a negative shock. Thus, the firms continue to follow their peer firms' ETR.

References

Admati, A. R., Pfleiderer, P., & Zechner, J. (1994). *Large Shareholder Activism, Risk Sharing, and Financial Market Equilibrium.* The Journal of Political Economy, 102(6), 1097-1130.

Armstrong, C. S., Glaeser, S., & Kepler, J. D. (2019). *Strategic reactions in corporate tax planning.* Journal of Accounting & Economics, 68(1), 101232.

Baltagi, B., Song, S., & Koh, W. (2003). *Testing Panel Data Regression Models with Spatial Error Correlation.* Journal of Econometrics 117(1):123-150.

Baskaran, T. (2015). *Tax Mimicking in the Short- and Long-Run: Evidence from German Reunification.* University of Goettingen, Department of Economics.

Beer, S., & Loeprick, J. (2015). *Profit shifting: Drivers of transfer (mis)pricing and the potential of countermeasures.* International Tax and Public Finance, 22(3), 426-451.

Betker, B. L. (1995). *An Empirical Examination of Prepackaged Bankruptcy.* Financial Management, 24(1), 3-18.

Bird, A., & Karolyi, S. A. (2017). *Governance and Taxes: Evidence from Regression Discontinuity (Retracted).* The Accounting Review, 92(1), 29-50.

Bird, A., & Karolyi, S. A. (2018). *Response to Alex Young.* Response to Alex Young. Econ Journal Watch, 15(1), 49.

Bird, A., Edwards, A., & Ruchti, T. G. (2018). *Taxes and Peer Effects.* The Accounting Review, 93(5), 97-117.

Brown, J. L., & Drake, K. D. (2014). *Network Ties Among Low-Tax Firms.* The Accounting Review, 89(2), 483-510.

Carpenter, M. A., Li, M., & Jiang, H. (2012). *Social Network Research in Organizational Contexts.* Journal of Management, 38(4), 1328-1361.

Cen, L., Maydew, E. L., Zhang, L., & Zuo, L. (2017). *Customer–supplier relationships and corporate tax avoidance.* Journal of Financial Economics, 123(2), 377-394.

Chen, S., Chen, X., Cheng, Q., & Shevlin, T. (2010). *Are family firms more tax aggressive than non-family firms?* Journal of Financial Economics, 95(1), 41-61.

Chen, S., Huang, Y., Li, N., & Shevlin, T. (2019). *How does quasi-indexer ownership affect corporate tax planning?* Journal of Accounting & Economics, 67(2-3), 278-296.

Chen, X., Harford, J., & Li, K. (2007). *Monitoring: Which institutions matter?* Journal of Financial Economics, 86(2), 279-305.

Cheng, C. A., Sun, Z., & Xie, J. (2018). *Common Equity Blockholders and Diffusion of Tax Avoidance.*

Cheng, C. S., Huang, H. H., Li, Y., & Stanfield, J. (2012). *The Effect of Hedge Fund Activism on Corporate Tax Avoidance.* The Accounting Review, 87(5), 1493-1526.

Clausing, K. A. (2009). *Multinational Firm Tax Avoidance and Tax Policy.* National Tax Journal, 62(4), 703-725.

Collischon, M., & Eberl, A. (2020). *Let's Talk About Fixed Effects: Let's Talk About All the Good Things and the Bad Things.* Kölner Zeitschrift Für Soziologie Und Sozialpsychologie, 72(2), 289-299.

Croissant, Y., & Millo, G. (2008). *Panel Data Econometrics in R: The plm Package.* Journal of Statistical Software 27(1):1–43.

Cronqvist, H., & Fahlenbrach, R. (2009). *Large Shareholders and Corporate Policies.* The Review of Financial Studies, 22(10), 3941-3976.

De Simone, L., Nickerson, J., Seidman, J., & Stomberg, B. (2020). *How Reliably Do Empirical Tests Identify Tax Avoidance?* Contemporary Accounting Research, 37(3), 1536-1561.

Desai, M. A., & Dharmapala, D. (2009). *CORPORATE TAX AVOIDANCE AND FIRM VALUE.* The Review of Economics and Statistics, 91(3), 537-546.

Dobrzynski, J. H. (1993). *Corporate Boards May Finally Be Shaping Up.* Bloomberg.

Dyreng, S. D., & Lindsey, B. P. (2009). *Using Financial Accounting Data to Examine the Effect of Foreign Operations Located in Tax Havens and Other Countries on U. S. Multinational Firms' Tax Rates.* Journal of Accounting Research, 47(5), 1283-1316.

Dyreng, S. D., Hanlon, M., Maydew, E. L., & Thornock, J. R. (2017). *Changes in corporate effective tax rates over the past 25 years.* Journal of Financial Economics, 124(3), 441-463.

Easterbrook, F. H., & Fischel, D. R. (1982). *Corporate Control Transactions.* The Yale Law Journal, 91(4), 698-737.

Edmans, A. (2009). *Blockholder Trading, Market Efficiency, and Managerial Myopia.* The Journal of Finance (New York), 64(6), 2481-2513.

Engelberg, J., Gao, P., & Parsons, C. A. (2012). *Friends with money.* Journal of Financial Economics, 103(1), 169-188.

Grossman, S. J., & Hart, O. D. (1980). *Takeover Bids, The Free-Rider Problem, and the Theory of the Corporation.* The Bell Journal of Economics, 11(1), 42-64.

Gulati, R., & Westphal, J. D. (1999). *Cooperative or Controlling? The Effects of CEO-Board Relations and the Content of Interlocks on the Formation of Joint Ventures.* Administrative Science Quarterly, 44(3), 473-506.

Gunaasih, S. A. (2021). *The Profitability, Leverage, and Company Size of the IDX80 Index on Tax Avoidance in Indonesia Stock Exchange.* Review of Integrative Business and Economics Research, 10, 106-113.

Hanlon, M., & Heitzman, S. (2010). *A review of tax research.* Journal of Accounting & Economics, 50(2), 127-178.

Hansen, R. G., & Lott, J. R. (1996). *Externalities and Corporate Objectives in a World with Diversified Shareholder/Consumers.* Journal of Financial and Quantitative Analysis, 31(1), 43-68.

Heitzman, S. M., & Ogneva, M. (2019). *Industry Tax Planning and Stock Returns.* The Accounting Review, 94(5), 219-246.

Henry, E., & Sansing, R. C. (2018). *Corporate Tax Avoidance: Data Truncation and Loss Firms.* Review of Accounting Studies, Forthcoming.

Hill, R. C., Griffiths, W. E., & Lim, G. C. (2018). *Principles of Econometrics, 5th Edition.* Wiley.

Kaplan, S. N., & Strömberg, P. (2009). *Leveraged Buyouts and Private Equity.* The Journal of Economic Perspectives, 23(1), 121-146.

Karjalainen, J., Kasanen, E., Kinnunen, J., & Niskanen, J. (2020). *Dividends and tax avoidance as drivers of earnings management: Evidence from dividend-paying private SMEs in Finland.* Journal of Small Business Management.

Karkinsky, T., & Riedel, N. (2012). *Corporate taxation and the choice of patent location within multinational firms.* Journal of International Economics, 88(1), 176-185.

Kartikaningdyah, E. (2019). *The Effect of Firm Size, ROA and Executive Character on Tax Avoidance.* Atlantis Press.

Khan, M., Srinivasan, S., & Tan, L. (2017). *Institutional Ownership and Corporate Tax Avoidance: New Evidence.* The Accounting Review, 92(2), 101-122.

Khurana, I. K., & Moser, W. J. (2013). *Institutional Shareholders' Investment Horizons and Tax Avoidance.* The Journal of the American Taxation Association, 35(1), 111-134.

Koh, P.-S., & Reeb, D. M. (2015). *Missing R&D.* Journal of Accounting & Economics, 60(1), 73-94.

Kubick, T. R., Lynch, D. P., Mayberry, M. A., & Omer, T. C. (2015). *Product Market Power and Tax Avoidance: Market Leaders, Mimicking Strategies, and Stock Returns.* . The Accounting Review, 90(2), 675-702.

Laplante, S. K., Skaife, H. A., Swenson, L. A., & Wangerin, D. D. (2019). *Limits of tax regulation: Evidence from strategic R&D classification and the R&D tax credit.* Journal of Accounting and Public Policy, 38(2), 89-105.

Leary, M. T., & Roberts, M. R. (2014). *Do Peer Firms Affect Corporate Financial Policy?* The Journal of Finance (New York), 69(1), 139-178.

Leipälä, A. (2017). *Do institutional shareholders impact corporate tax avoidance? (Master's thesis).* Aalto University.

Lopes da Fonseca, M. (2017). *Tax Mimicking in Local Business Taxation: Quasi-Experimental Evidence from Portugal.* Social Science Research Network.

Manski, C. F. (1993). *Identification of Endogenous Social Effects: The Reflection Problem.* The Review of Economic Studies, 60(3), 531-542.

Monks, R. A., & Minow, N. (2011). *Corporate Governance, 5th edition.* Wiley.

Neubig, T., & Wunsch-Vincent, S. (2018). *Tax distortions in cross-border flows of intangible assets.* International Journal of Innovation Studies, 2 (3), 101-121.

Parrino, R., Sias, R. W., & Starks, L. T. (2003). *Voting with their feet: Institutional ownership changes around forced CEO turnover.* Journal of Financial Economics, 68(1), 3-46.

Peters, R. H., & Taylor, L. A. (2017). *Intangible capital and the investment-q relation.* Journal of Financial Economics, 123(2), 251-272.

Roald, A. K., & Roti, T. (2015). *Tax Planning in Norwegian Private Equity-backed Companies (Master's thesis).* NHH.

Rubin, A. (2006). *Diversification and Corporate Decisions.* Corporate Ownership and Control 3 (3), 209-212.

Schmidberger, M., Morgan, M., Eddelbuettel, D., Yu, H., Tierney, L., & Mansmann, U. (2009). *State of the Art in Parallel Computing with R.* State of the Art in Parallel Computing with R». Journal of Statistical Software 31(1):1–27.

Shackelford, D. A., & Shevlin, T. (2001). *Empirical tax research in accounting.* Journal of Accounting & Economics, 31(1), 321-387.

Shevlin, T., Lim, C. Y., Lee, J., & Goh, B. W. (2016). *The Effect of Corporate Tax Avoidance on the Cost of Equity.* The Accounting Review, 91(6), 1647-1670.

Streitferdt, F. G. (2013). *The Valuation of Tax Loss Carryforwards.* Journal of Business Valuation and Economic Loss Analysis, 8(1), 163-184.

Wooldridge, J. (2012). *Introductory Econometrics: A Modern Approach.* South-Western College Publishing.

Yan, X., & Zhang, Z. (2009). *Institutional Investors and Equity Returns: Are Short-term Institutions Better Informed?* The Review of Financial Studies, 22(2), 893-924.

Tables

Table 1.1: Summary statistics

Table 1 presents summary statistics for our sample in section 8.2. It reports descriptive statistics of the variables, including our primary dependent variable: effective tax rate (*GAAP_ETR and Cash ETR*) and our primary independent variable of peer firm's effective tax rate (*PEER_ETR*). Control variables including intangible assets (*INTAN)*, institutional ownership (*INST*), market-to-book ratio (*MB*), loss carryforward (*NOL*), change in loss carryforward (*DIFFNOL),* property, plant, and equipment (*PPE)*, profitability (*ROA*), market capitalization (*MCAP*), foreign income and loss (*F_INC*), equity income earnings (*EQINC*), leverage ratio (*LEV*), R&D (*R_D*), blank R&D (*R_D_BLANK*), Herfindahl-Hirschman Index (*HHI*). Appendix A provides definitions of variables. The GAAP ETR sample consists of 23,303 firm-year observations from the year 1999 to 2016, while the cash ETR sample

consists of 23,015 firm-year observations. The tables show statistics after winsorizing the control variables; *INTAN*, *INST*, *MB, DIFFNOL*, *PPE*, *ROA*, *MKTCAP*, *F_INC*, *EQINC*, *LEV*, and *R_D*, to the 1st and 99th percentiles.

Panel A: GAAP ETR sample

Statistic	N	Mean	Q1	Median	Q3	St. Dev.
GAAP_ETR	23,603	0.327	0.281	0.349	0.384	0.123
peer_etr	23,603	0.321	0.303	0.321	0.341	0.039
intan	23,603	0.180	0.015	0.117	0.291	0.190
inst	23,603	0.663	0.483	0.710	0.866	0.254
mb	23,603	3.101	1.415	2.231	3.656	4.331
nol	23,603	0.452	0	0	1	0.498
diffnol	23,603	-0.001	0.000	0.000	0.000	0.061
ppe	23,603	0.249	0.085	0.182	0.348	0.215
roa	23,603	0.105	0.052	0.089	0.141	0.076
mktcap	23,603	4,102.130	245.453	821.226	2,805.987	9,980.648
f_inc	23,603	0.018	0	0	0.03	0.033
eqinc	23,603	0.001	0	0	0	0.003
lev	23,603	0.171	0.001	0.133	0.276	0.179
r_d	23,603	0.049	0.004	0.02	0.1	0.063
r_d_blank	23,603	0.359	0	0	1	0.480
HHI	23,603	978.521	499.382	655.468	1,162.065	816.045

Panel B: Cash ETR sample

Statistic	N	Mean	Q1	Median	Q3	St. Dev.
CASH_ETR	23,015	0.255	0.122	0.250	0.351	0.173
peer_etr	23,015	0.250	0.223	0.251	0.274	0.052
intan	23,015	0.182	0.017	0.120	0.295	0.191
inst	23,015	0.666	0.488	0.713	0.869	0.253
mb	23,015	3.121	1.443	2.267	3.693	4.342
nol	23,015	0.460	0	0	1	0.498
diffnol	23,015	-0.001	0.000	0.000	0.000	0.062
ppe	23,015	0.244	0.083	0.178	0.339	0.212
roa	23,015	0.107	0.054	0.090	0.142	0.074
mktcap	23,015	4,162.405	251.975	836.444	2,854.822	10,082.350
f_inc	23,015	0.019	0	0	0.03	0.033
eqinc	23,015	0.001	0	0	0	0.003

lev	23,015	0.169	0.001	0.129	0.274	0.179
r_d	23,015	0.048	0.005	0.02	0.1	0.060
r_d_blank	23,015	0.351	0	0	1	0.477
HHI	23,015	969.118	498.898	651.794	1,149.243	804.105

Table 1.2: SIC statistics of peer groups

A peer group is identified by looking at the unique CIBs in our sample and their corresponding portfolio. The average of CIBs with shares in firms with the same SIC is calculated by counting the number of firms in the portfolio and the number of those firms' unique SIC's. First, the calculation is done yearly, and then we calculate the mean across all the years in the sample to get the numbers below. This table represents statistics for all U.S. incorporated firms which do not operate in the unity or financial industry before any filtering (e.g., negative pretax-income). For results in our filtered sample, see section 10.3.

Number of unique CIB's	2,478
Average of CIBs with shares in firms sharing the same 4-digit SIC	5.74%
Average of CIBs with shares in firms sharing the same 3-digit SIC	8,21%
Average of CIBs with shares in firms sharing the same 2-digit SIC	11,62%
Average of CIBs with shares in firms sharing the same 1-digit SIC	19,83%

Table 2: Baseline regression

Table for section 8.5.

This table presents the OLS regression results of associations between firm's tax avoidance and their CIB's peer for the pooled sample. The dependent variable is the focal firm's effective tax rate (*GAAP_ETR and CASH_ETR*), an independent variable is peer firms' effective tax rate (*PEER_ETR*). For a focal firm *I* in year *t*, we use its holding information in year *t-1* to determine its CIB peers. Appendix A provides definitions of control variables. We control for both year and industry fixed effects. Standard errors clustered at the firm level. *, **, *** indicates significance level of 10%, 5% and 1% respectively.

Panel A: GAAP ETR

Baseline regression analysis

	Dependent variable: Focal firm's GAAP ETR (t)		
	(1)	(2)	(3)
peer_etr (t-1)	0.330***	0.202***	0.187***
	(0.029)	(0.027)	(0.026)
intan (t-1)		0.020***	0.021***
		(0.006)	(0.007)
inst (t-1)		0.004	0.005
		(0.005)	(0.005)
lev (t-1)		0.013*	0.012

		(0.007)	(0.007)
mb (t-1)		0.00004	-0.00001
		(0.0002)	(0.0002)
nol (t-1)		-0.008***	-0.010***
		(0.002)	(0.002)
diffnol (t-1)		-0.021	-0.026
		(0.024)	(0.024)
ppe (t-1)		0.001	0.007
		(0.007)	(0.008)
roa (t-1)		0.154***	0.153***
		(0.015)	(0.015)
log_mktcap (t-1)		-0.004***	-0.005***
		(0.001)	(0.001)
f_inc (t-1)		-0.655***	-0.556***
		(0.046)	(0.046)
eqinc (t-1)		-0.836**	-0.701**
		(0.351)	(0.344)
r_d (t-1)		-0.207***	-0.179***
		(0.023)	(0.027)
r_d_blank (t-1)		0.032***	0.019***
		(0.002)	(0.003)
HHI (t-1)		0.00000*	-0.00001**
		(0.00000)	(0.00000)
Constant	0.226***		
	(0.009)		
Control variables	No	Yes	Yes
Year fixed effects	No	Yes	Yes
Industry fixed effects	No	No	Yes
Observations	17,003	17,003	17,003
R^2	0.012	0.120	0.158
Significance levels		*p<0.1; **p<0.05; ***p<0.01	

Panel B: Cash ETR

Baseline regression analysis

	Dependent variable: Focal firm's cash ETR (t)		
	(1)	(2)	(3)

peer_etr (t-1)	0.247***	0.224***	0.197***
	(0.029)	(0.028)	(0.027)
intan (t-1)		0.022**	0.039***
		(0.011)	(0.012)
inst (t-1)		-0.008	-0.007
		(0.008)	(0.008)
lev (t-1)		-0.034***	-0.045***
		(0.012)	(0.012)
mb (t-1)		-0.002***	-0.002***
		(0.0004)	(0.0004)
nol (t-1)		-0.030***	-0.028***
		(0.004)	(0.003)
diffnol (t-1)		0.082***	0.076***
		(0.030)	(0.029)
ppe (t-1)		-0.081***	-0.047***
		(0.011)	(0.013)
roa (t-1)		0.326***	0.299***
		(0.023)	(0.023)
log_mktcap (t-1)		0.0002	0.0001
		(0.001)	(0.001)
f_inc (t-1)		-0.084	0.004
		(0.062)	(0.063)
eqinc (t-1)		-0.309	-0.264
		(0.507)	(0.489)
r_d (t-1)		-0.383***	-0.271***
		(0.033)	(0.038)
r_d_blank (t-1)		0.022***	0.030***
		(0.004)	(0.005)
HHI (t-1)		0.00001***	0.00001
		(0.00000)	(0.00000)
Constant	0.204***		
	(0.008)		
Control variables	No	Yes	Yes
Year fixed effects	No	Yes	Yes
Industry fixed effects	No	No	Yes
Observations	16,638	16,638	16,638
R^2	0.006	0.072	0.102
Significance levels		*p<0.1; **p<0.05; ***p<0.01	

Table 3: Long-term vs. short-term blockholders in focal firms

Regression for section 8.6. Dependent variable is focal firm's ETR in year *t*. Both independent and control variables are in year *t-1*. Control variables abbreviations are explained in Appendix A.

Panel A: GAAP ETR

Analysis: Long-term vs. short-term CIB's

	Dependent variable: Focal firm's GAAP ETR (t)
peer_etr_long (t-1)	0.085*
	(0.046)
peer_etr_short (t-1)	0.076**
	(0.035)
intan (t-1)	0.013
	(0.012)
inst (t-1)	0.014
	(0.009)
lev (t-1)	-0.004
	(0.014)
mb (t-1)	0.0002
	(0.0004)
nol (t-1)	-0.001
	(0.003)
diffnol (t-1)	0.016
	(0.036)
ppe (t-1)	0.001
	(0.014)
roa (t-1)	0.180***
	(0.024)
log_mktcap (t-1)	-0.006***
	(0.001)
f_inc (t-1)	-0.619***
	(0.064)
eqinc (t-1)	-1.432***

	(0.505)
r_d (t-1)	-0.154***
	(0.047)
r_d_blank (t-1)	0.024***
	(0.005)
HHI (t-1)	-0.00000
	(0.00000)
Control variables	Yes
Industry fixed effects	Yes
Year fixed effects	Yes
Long-term definition:	Blockholder for 4+ subsequent years
Observations	5,621
R^2	0.197
Significance levels	*p<0.1; **p<0.05; ***p<0.01

Panel B: Cash ETR

Analysis: Long-term vs. short-term CIB's

	Dependent variable: Focal firm's cash ETR (t)
peer_etr_long (t-1)	0.046
	(0.043)
peer_etr_short (t-1)	0.186***
	(0.044)
intan (t-1)	0.033*
	(0.018)
inst (t-1)	-0.006
	(0.017)
lev (t-1)	-0.061***
	(0.019)
mb (t-1)	-0.002***
	(0.001)
nol (t-1)	-0.014***
	(0.005)
diffnol (t-1)	0.172***
	(0.053)
ppe (t-1)	-0.060***
	(0.023)

roa (t-1)	0.309***
	(0.038)
log_mktcap (t-1)	-0.002
	(0.002)
f_inc (t-1)	-0.123
	(0.088)
eqinc (t-1)	-1.035
	(0.646)
r_d (t-1)	-0.227***
	(0.061)
r_d_blank (t-1)	0.028***
	(0.007)
HHI (t-1)	0.00001
	(0.00001)
Control variables	Yes
Industry fixed effects	Yes
Year fixed effects	Yes
Long-term definition:	Blockholder for 4+ subsequent years
Observations	5,593
R^2	0.118
Significance levels	*p<0.1; **p<0.05; ***p<0.01

Table 4: Long-term vs. short-term blockholder in peer firms

Regression for section 8.7. Appendix A shows detailed explanations of control variables. We control for industry and firm fixed effects, and standard errors are clustered at the firm level. *, **, *** indicate significance level of 10%, 5% and 1% respectively.

Panel A: GAAP ETR

Analysis: Long-term vs. short-term sub-portfolio's

	Dependent variable: Focal firm's GAAP ETR (t)
peer_etr_long-term_portfolio (t-1)	0.065***
	(0.023)
peer_etr_short-term_portfolio (t-1)	0.108***
	(0.031)
intan (t-1)	0.019**
	(0.008)

inst (t-1)	0.012**
	(0.006)
lev (t-1)	0.005
	(0.009)
mb (t-1)	0.0001
	(0.0002)
nol (t-1)	-0.007***
	(0.002)
diffnol (t-1)	-0.028
	(0.027)
ppe (t-1)	0.006
	(0.009)
roa (t-1)	0.176***
	(0.016)
log_mktcap (t-1)	-0.005***
	(0.001)
f_inc (t-1)	-0.616***
	(0.048)
eqinc (t-1)	-0.981**
	(0.396)
r_d (t-1)	-0.200***
	(0.034)
r_d_blank (t-1)	0.020***
	(0.003)
HHI (t-1)	-0.00000
	(0.00000)
Control variables	Yes
Industry fixed effects	Yes
Year fixed effects	Yes
Long-term definition:	Blockholder for 4+ subsequent years
Observations	13,061
R^2	0.182
Significance levels	*p<0.1; **p<0.05; ***p<0.01

Panel B: Cash ETR

Analysis: Long-term vs. short-term sub-portfolio's

	Dependent variable: Focal firm's cash ETR (t)

peer_etr_long-term_portfolio (t-1)	0.066**
	(0.027)
peer_etr_short-term_portfolio (t-1)	0.193***
	(0.032)
intan (t-1)	0.040***
	(0.013)
inst (t-1)	-0.003
	(0.009)
lev (t-1)	-0.053***
	(0.013)
mb (t-1)	-0.002***
	(0.0004)
nol (t-1)	-0.023***
	(0.004)
diffnol (t-1)	0.089***
	(0.032)
ppe (t-1)	-0.034**
	(0.015)
roa (t-1)	0.304***
	(0.025)
log_mktcap (t-1)	0.001
	(0.001)
f_inc (t-1)	-0.064
	(0.067)
eqinc (t-1)	-0.286
	(0.524)
r_d (t-1)	-0.304***
	(0.045)
r_d_blank (t-1)	0.032***
	(0.005)
HHI (t-1)	0.00001
	(0.00001)
Control variables	Yes
Industry fixed effects	Yes
Year fixed effects	Yes
Long-term definition:	Blockholder for 4+ subsequent years
Observations	12,939
R^2	0.107

Significance levels	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 5: Largest CIB regression

Regression for section 9.1. Appendix A shows detailed explanations of control variables. We control for industry and firm fixed effects, and standard errors are clustered at the firm level. *, **, *** indicate significance level of 10%, 5% and 1% respectively.

Panel A: GAAP ETR

	Largest CIB's regression
	Dependent variable: Focal firm's GAAP ETR (t)
peer_etr_largest (t-1)	0.076^{***}
	(0.022)
peer_etr_rest (t-1)	0.163^{***}
	(0.029)
intan (t-1)	0.020^{**}
	(0.008)
inst (t-1)	0.003
	(0.007)
lev (t-1)	0.005
	(0.009)
mb (t-1)	0.0001
	(0.0003)
nol (t-1)	-0.009^{***}
	(0.002)
diffnol (t-1)	-0.024
	(0.030)
ppe (t-1)	0.009
	(0.009)
roa (t-1)	0.163^{***}
	(0.017)
log_mktcap (t-1)	-0.005^{***}
	(0.001)
f_inc (t-1)	-0.587^{***}
	(0.054)
eqinc (t-1)	-0.803^{**}
	(0.377)
r_d (t-1)	-0.150^{***}

	(0.032)
r_d_blank (t-1)	0.019***
	(0.003)
HHI (t-1)	-0.00000*
	(0.00000)
Control variables	Yes
Industry fixed effects	Yes
Year fixed effects	Yes
Observations	12,709
R^2	0.174
Significance levels	*p<0.1; **p<0.05; ***p<0.01

Panel B: Cash ETR

Largest CIB's regression

	Dependent variable: Focal firm's cash ETR (t)
peer_etr_largest (t-1)	0.117***
	(0.025)
peer_etr_rest (t-1)	0.174***
	(0.031)
intan (t-1)	0.014
	(0.013)
inst (t-1)	-0.018*
	(0.011)
lev (t-1)	-0.031**
	(0.014)
mb (t-1)	-0.002***
	(0.0004)
nol (t-1)	-0.022***
	(0.004)
diffnol (t-1)	0.169***
	(0.041)
ppe (t-1)	-0.078***
	(0.016)
roa (t-1)	0.289***
	(0.025)
log_mktcap (t-1)	-0.003*
	(0.002)

f_inc (t-1)	-0.017
	(0.072)
eqinc (t-1)	0.152
	(0.502)
r_d (t-1)	-0.205***
	(0.044)
r_d_blank (t-1)	0.028***
	(0.005)
HHI (t-1)	0.00001*
	(0.00000)
Control variables	Yes
Industry fixed effects	Yes
Year fixed effects	Yes
Observations	11,477
R^2	0.109
Significance levels	*p<0.1; **p<0.05; ***p<0.01

Table 6: Firm fixed effects regression model

Regression for section 10.1.

This table presents the OLS regression results of associations between the firm's tax avoidance and their CIB's peer for the pooled sample with firm fixed effects. The dependent variable is the focal firm's effective tax rate (*GAAP_ETR* and *Cash_ETR*). For a focal firm *i* in year *t*, we determine its CIB peers using holding information in year *t-1*. *, **, *** indicate significance level of 10%, 5% and 1% respectively.

Panel A: GAAP ETR

Firm fixed effects regression	
	Dependent variable: Focal firm's GAAP ETR (t)
peer_etr (t-1)	0.149***
	(0.029)
intan (t-1)	0.016
	(0.010)
inst (t-1)	0.003
	(0.007)
lev (t-1)	0.009
	(0.010)
mb (t-1)	0.0002
	(0.0002)

nol (t-1)	-0.007**
	(0.003)
diffnol (t-1)	-0.002
	(0.025)
ppe (t-1)	-0.004
	(0.010)
roa (t-1)	0.149***
	(0.019)
log_mktcap (t-1)	-0.005***
	(0.001)
f_inc (t-1)	-0.508***
	(0.067)
eqinc (t-1)	-1.213***
	(0.427)
r_d (t-1)	-0.248***
	(0.033)
r_d_blank (t-1)	0.031***
	(0.004)
HHI (t-1)	0.00000
	(0.00000)
Control variables	Yes
Firm fixed effects	Yes
Year fixed effects	Yes
Observations	17,003
R^2	0.058
Significance levels	*p<0.1; **p<0.05; ***p<0.01

Panel B: Cash ETR

Firm fixed effects regression

	Dependent variable: Focal firm's cash ETR (t)
peer_etr (t-1)	0.173***
	(0.029)
intan (t-1)	0.049***
	(0.015)
inst (t-1)	0.002
	(0.011)
lev (t-1)	-0.023

	(0.015)
mb (t-1)	-0.002***
	(0.0004)
nol (t-1)	-0.026***
	(0.004)
diffnol (t-1)	0.025
	(0.029)
ppe (t-1)	-0.075***
	(0.015)
roa (t-1)	0.350***
	(0.030)
log_mktcap (t-1)	-0.002
	(0.002)
f_inc (t-1)	0.050
	(0.077)
eqinc (t-1)	-1.015
	(0.696)
r_d (t-1)	-0.335***
	(0.044)
r_d_blank (t-1)	0.015**
	(0.006)
HHI (t-1)	0.00001**
	(0.00000)
Control variables	Yes
Firm fixed effects	Yes
Year fixed effects	Yes
Observations	16,638
R^2	0.045
Significance levels	*p<0.1; **p<0.05; ***p<0.01

Table 7: SIC restrictions

Regressions for section 10.2. Peer firms with the same SIC as focal firms are excluded in the peer ETR calculation. Dependent variable is focal firm's ETR in year *t*. All other variables are in year *t-1*. Control variables abbreviations are explained in Appendix A. Year and industry fixed effects are controlled for, and standard errors are clustered at the firm level. *,**, and *** indicate significance level of 10%, 5% and 1% respectively.

Panel A: GAAP ETR

Panel A.1: 4-digit SIC restriction

4-digit SIC restriction	
	Dependent variable: Focal firm's GAAP ETR (t)
peer_etr (t-1)	0.164***
	(0.027)
intan (t-1)	0.021***
	(0.007)
inst (t-1)	0.004
	(0.005)
lev (t-1)	0.012
	(0.007)
mb (t-1)	-0.00001
	(0.0002)
nol (t-1)	-0.010***
	(0.002)
diffnol (t-1)	-0.026
	(0.023)
ppe (t-1)	0.007
	(0.008)
roa (t-1)	0.154***
	(0.014)
log_mktcap (t-1)	-0.005***
	(0.001)
f_inc (t-1)	-0.558***
	(0.045)
eqinc (t-1)	-0.704**
	(0.342)
r_d (t-1)	-0.179***
	(0.028)
r_d_blank (t-1)	0.019***
	(0.003)
HHI (t-1)	-0.00001**
	(0.00000)
Control variables	Yes
Industry fixed effects	Yes
Year fixed effects	Yes
Observations	16,989

R^2	0.157
Significance levels	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Panel A.2: 3-digit SIC restriction

3-digit SIC restriction

	Dependent variable: Focal firm's GAAP ETR (t)
peer_etr (t-1)	0.158***
	(0.028)
intan (t-1)	0.021***
	(0.007)
inst (t-1)	0.004
	(0.005)
lev (t-1)	0.012
	(0.008)
mb (t-1)	-0.00000
	(0.0002)
nol (t-1)	-0.010***
	(0.002)
diffnol (t-1)	-0.025
	(0.024)
ppe (t-1)	0.007
	(0.008)
roa (t-1)	0.153***
	(0.014)
log_mktcap (t-1)	-0.005***
	(0.001)
f_inc (t-1)	-0.557***
	(0.046)
eqinc (t-1)	-0.690*
	(0.352)
r_d (t-1)	-0.180***
	(0.028)
r_d_blank (t-1)	0.019***
	(0.003)
HHI (t-1)	-0.00001**
	(0.00000)
Control variables	Yes

Industry fixed effects	Yes
Year fixed effects	Yes
Observations	16,983
R^2	0.157
Significance levels	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Panel A.3: 2-digit SIC restriction

2-digit SIC restriction

	Dependent variable: Focal firm's GAAP ETR (t)
peer_etr (t-1)	0.165^{***}
	(0.029)
intan (t-1)	0.020^{***}
	(0.007)
inst (t-1)	0.004
	(0.005)
lev (t-1)	0.012
	(0.007)
mb (t-1)	-0.00001
	(0.0002)
nol (t-1)	-0.010^{***}
	(0.002)
diffnol (t-1)	-0.026
	(0.024)
ppe (t-1)	0.007
	(0.008)
roa (t-1)	0.155^{***}
	(0.014)
log_mktcap (t-1)	-0.005^{***}
	(0.001)
f_inc (t-1)	-0.560^{***}
	(0.045)
eqinc (t-1)	-0.766^{**}
	(0.343)
r_d (t-1)	-0.184^{***}
	(0.028)
r_d_blank (t-1)	0.019^{***}
	(0.003)

HHI (t-1)	-0.00001**
	(0.00000)
Control variables	Yes
Industry fixed effects	Yes
Year fixed effects	Yes
Observations	16,971
R^2	0.158
Significance levels	*p<0.1; **p<0.05; ***p<0.01

Panel B: Cash ETR

Panel B.1: 4-digit SIC restriction

	4-digit SIC restriction
	Dependent variable: Focal firm's cash ETR (t)
peer_etr (t-1)	0.187***
	(0.028)
intan (t-1)	0.039***
	(0.012)
inst (t-1)	-0.007
	(0.008)
lev (t-1)	-0.045***
	(0.012)
mb (t-1)	-0.002***
	(0.0003)
nol (t-1)	-0.028***
	(0.003)
diffnol (t-1)	0.076***
	(0.029)
ppe (t-1)	-0.046***
	(0.013)
roa (t-1)	0.300***
	(0.022)
log_mktcap (t-1)	0.0001
	(0.001)
f_inc (t-1)	0.005
	(0.063)
eqinc (t-1)	-0.266

	(0.487)
r_d (t-1)	-0.274***
	(0.039)
r_d_blank (t-1)	0.030***
	(0.005)
HHI (t-1)	0.00000
	(0.00000)
Control variables	Yes
Industry fixed effects	Yes
Year fixed effects	Yes
Observations	16,626
R^2	0.101
Significance levels	*p<0.1; **p<0.05; ***p<0.01

Panel B.2: 3-digit SIC restriction

3-digit SIC restriction	
	Dependent variable: Focal firm's cash ETR (t)
peer_etr (t-1)	0.178***
	(0.028)
intan (t-1)	0.040***
	(0.011)
inst (t-1)	-0.007
	(0.008)
lev (t-1)	-0.045***
	(0.012)
mb (t-1)	-0.002***
	(0.0003)
nol (t-1)	-0.029***
	(0.003)
diffnol (t-1)	0.079**
	(0.031)
ppe (t-1)	-0.046***
	(0.014)
roa (t-1)	0.299***
	(0.022)
log_mktcap (t-1)	0.0001
	(0.001)

f_inc (t-1)	0.008
	(0.061)
eqinc (t-1)	-0.265
	(0.500)
r_d (t-1)	-0.270***
	(0.039)
r_d_blank (t-1)	0.030***
	(0.005)
HHI (t-1)	0.00000
	(0.00000)
Control variables	Yes
Industry fixed effects	Yes
Year fixed effects	Yes
Observations	16,618
R^2	0.101
Significance levels	*p<0.1; **p<0.05; ***p<0.01

Panel B.3: 2-digit SIC restriction

	2-digit SIC restriction
	Dependent variable: Focal firm's cash ETR (t)
peer_etr (t-1)	0.174***
	(0.027)
intan (t-1)	0.040***
	(0.012)
inst (t-1)	-0.007
	(0.008)
lev (t-1)	-0.045***
	(0.012)
mb (t-1)	-0.002***
	(0.0003)
nol (t-1)	-0.028***
	(0.003)
diffnol (t-1)	0.078***
	(0.029)
ppe (t-1)	-0.046***
	(0.014)
roa (t-1)	0.302***

	(0.022)
log_mktcap (t-1)	0.00004
	(0.001)
f_inc (t-1)	0.010
	(0.062)
eqinc (t-1)	-0.311
	(0.466)
r_d (t-1)	-0.269***
	(0.040)
r_d_blank (t-1)	0.030***
	(0.005)
HHI (t-1)	0.00000
	(0.00000)
Control variables	Yes
Industry fixed effects	Yes
Year fixed effects	Yes
Observations	16,602
R^2	0.101
Significance levels	*p<0.1; **p<0.05; ***p<0.01

Table 8: Event study on shocks in peer ETR's

Regression for section 10.3.

The table displays the results of peer effect on a firm's tax avoidance using change in tax rate level as exogenous events. Year *t* is the event year when the tax rate is either increased or decreased. Independent variables are *one_year_after_negative_shock* and *one_year_after_the_positive_shock.* They capture the effect of a decrease/increase in peer's tax rate on a focal firm in the year following the shock. Besides the control variables, we also control whether the effect is inherent before and persist after the shock. These control variables are *1-4+ years before positive/negative shocks* and *1-5+years after positive/negative shocks.* Columns (1) and (2) show the two different magnitudes of tax shocks, with a change between 5% and 10% and a change of more than 10%, respectively. We control for year, firm, and industry fixed effects. Standard errors are clustered at the firm level. *, **, *** indicate significance at the 10%, 5% and 1% levels.

Panel A: GAAP_ETR

Event: Shock in GAAP peer ETR

	Dependent variable: Focal firm's GAAP ETR	
	(1)	(2)
5+ years before neg shock	0.021***	-0.0001

	(0.006)	(0.014)
4 years before neg shock	0.012*	-0.005
	(0.007)	(0.018)
3 years before neg shock	0.012*	-0.015
	(0.006)	(0.015)
2 years before neg shock	0.020***	-0.005
	(0.006)	(0.011)
Negative shock year	0.003	-0.004
	(0.005)	(0.011)
1 year after neg shock	0.007	-0.001
	(0.006)	(0.010)
2 years after neg shock	0.005	-0.034***
	(0.006)	(0.012)
3 years after neg shock	-0.009	-0.004
	(0.006)	(0.013)
4 years after neg shock	-0.010	-0.014
	(0.007)	(0.011)
5+ years after neg shock	-0.008	-0.019*
	(0.007)	(0.011)
5+ years before pos shock	0.006	-0.010
	(0.007)	(0.014)
4 years before pos shock	0.004	0.006
	(0.009)	(0.021)
3 years before pos shock	0.012	0.012
	(0.008)	(0.014)
2 years before pos shock	0.005	-0.0001
	(0.007)	(0.011)
Positive shock year	0.010*	0.010
	(0.005)	(0.009)
1 year after pos shock	0.017***	-0.0002
	(0.006)	(0.010)
2 years after pos shock	0.016**	-0.015
	(0.007)	(0.010)
3 years after pos shock	0.001	-0.006
	(0.007)	(0.011)
4 years after pos shock	0.001	0.016
	(0.007)	(0.015)
5+ years after pos shock	-0.002	-0.018

	(0.007)	(0.013)
intan	0.033***	0.033***
	(0.009)	(0.009)
inst	0.008	0.006
	(0.007)	(0.007)
lev	-0.008	-0.006
	(0.009)	(0.009)
mb	-0.0004	-0.0004
	(0.0003)	(0.0003)
nol	-0.020***	-0.020***
	(0.003)	(0.003)
diffnol	0.033	0.032
	(0.023)	(0.023)
ppe	-0.006	-0.006
	(0.009)	(0.009)
roa	-0.006	-0.003
	(0.017)	(0.017)
log_mktcap	-0.0002	-0.0003
	(0.001)	(0.001)
f_inc	-0.559***	-0.559***
	(0.056)	(0.057)
eqinc	-1.503***	-1.464***
	(0.418)	(0.418)
r_d	-0.257***	-0.261***
	(0.026)	(0.026)
r_d_blank	0.032***	0.032***
	(0.004)	(0.004)
HHI	0.00000	0.00000
	(0.00000)	(0.00000)
Control variables	Yes	Yes
Firm fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Shock	5% - 10%	10% ->
Observations	23,603	23,603
R^2	0.053	0.052
Significance levels		*p<0.1; **p<0.05; ***p<0.01

Panel B: Cash ETR

Event: Shock in cash peer ETR

	Dependent variable: Focal firm's cash ETR	
	(1)	(2)
5+ years before neg shock	-0.011	0.009
	(0.007)	(0.014)
4 years before neg shock	0.012*	0.007
	(0.007)	(0.014)
3 years before neg shock	0.011*	0.025*
	(0.006)	(0.013)
2 years before neg shock	0.017***	-0.003
	(0.006)	(0.010)
Negative shock year	0.008	-0.004
	(0.005)	(0.009)
1 year after neg shock	0.011*	-0.003
	(0.006)	(0.009)
2 years after neg shock	0.020***	0.030***
	(0.006)	(0.011)
3 years after neg shock	0.022***	0.026**
	(0.006)	(0.011)
4 years after neg shock	0.034***	0.028**
	(0.007)	(0.011)
5+ years after neg shock	0.027***	0.028**
	(0.007)	(0.011)
5+ years before pos shock	0.024***	0.017
	(0.007)	(0.014)
4 years before pos shock	0.007	-0.006
	(0.006)	(0.014)
3 years before pos shock	-0.012**	-0.001
	(0.006)	(0.013)
2 years before pos shock	-0.011**	-0.017
	(0.005)	(0.010)
Positive shock year	0.016***	0.016**
	(0.005)	(0.008)
1 year after pos shock	0.027***	0.017*
	(0.006)	(0.009)
2 years after pos shock	0.020***	0.024***

	(0.006)	(0.009)
3 years after pos shock	0.016**	0.013
	(0.006)	(0.011)
4 years after pos shock	0.022***	0.023**
	(0.008)	(0.012)
5+ years after pos shock	0.020***	0.014
	(0.007)	(0.010)
intan	0.048***	0.053***
	(0.013)	(0.013)
inst	-0.013	-0.002
	(0.010)	(0.010)
lev	-0.048***	-0.051***
	(0.013)	(0.013)
mb	-0.001***	-0.001***
	(0.0004)	(0.0004)
nol	-0.043***	-0.042***
	(0.004)	(0.004)
diffnol	0.090***	0.090***
	(0.024)	(0.024)
ppe	-0.070***	-0.065***
	(0.013)	(0.013)
roa	-0.185***	-0.175***
	(0.028)	(0.028)
log_mktcap	0.001	0.002
	(0.002)	(0.002)
f_inc	-0.093	-0.077
	(0.066)	(0.067)
eqinc	-1.487**	-1.525***
	(0.586)	(0.586)
r_d	-0.172***	-0.187***
	(0.038)	(0.038)
r_d_blank	0.015***	0.014***
	(0.005)	(0.005)
HHI	0.00001***	0.00001***
	(0.00000)	(0.00000)
Control variables	Yes	Yes
Firm fixed effects	Yes	Yes
Year fixed effects	Yes	Yes

Shock	5% - 10%	10% ->
Observations	23,015	23,015
R^2	0.039	0.034
Significance levels		$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 9: Investor level event study

Regression for section 10.4.

Control variables abbreviations are explained in Appendix A. Year and individual fixed effects are controlled for, and standard errors are clustered at the firm level. *,**, and *** indicate significance level of 10%, 5% and 1% respectively.

Panel A: GAAP ETR

Event: Shock in GAAP ETR

	Dependent variable: Blockholder's portfolio weight	
	(1)	(2)
5+ years before neg shock	0.003	-0.005
	(0.004)	(0.006)
4 years before neg shock	-0.003	-0.007
	(0.003)	(0.005)
3 years before neg shock	-0.0004	0.0003
	(0.002)	(0.005)
2 years before neg shock	0.001	-0.003
	(0.002)	(0.004)
Negative shock year	-0.002	0.002
	(0.002)	(0.003)
1 year after neg shock	-0.006**	0.002
	(0.002)	(0.004)
2 years after neg shock	-0.0002	0.003
	(0.003)	(0.005)
3 years after neg shock	-0.003	0.001
	(0.004)	(0.006)
4 years after neg shock	-0.004	0.004
	(0.004)	(0.006)
5+ years after neg shock	-0.008	0.012
	(0.005)	(0.008)

5+ years before pos shock	0.0003	-0.006
	(0.004)	(0.006)
4 years before pos shock	-0.002	-0.0004
	(0.003)	(0.005)
3 years before pos shock	-0.004	0.001
	(0.003)	(0.005)
2 years before pos shock	-0.002	0.004
	(0.002)	(0.004)
Positive shock year	-0.001	0.006**
	(0.002)	(0.003)
1 year after pos shock	0.003	0.001
	(0.003)	(0.004)
2 years after pos shock	0.003	0.008
	(0.003)	(0.005)
3 years after pos shock	-0.003	0.006
	(0.003)	(0.007)
4 years after pos shock	-0.001	-0.002
	(0.004)	(0.007)
5+ years after pos shock	-0.006	-0.003
	(0.006)	(0.007)
intan	-0.021**	-0.022**
	(0.009)	(0.009)
inst	-0.007	-0.006
	(0.007)	(0.007)
lev	0.018**	0.018**
	(0.008)	(0.008)
mb	0.0002	0.0002
	(0.0002)	(0.0002)
nol	-0.002	-0.002
	(0.002)	(0.002)
diffnol	-0.025	-0.021
	(0.027)	(0.027)
ppe	-0.016	-0.015
	(0.016)	(0.016)
roa	-0.001	-0.001
	(0.015)	(0.015)
log_mktcap	0.019***	0.019***
	(0.002)	(0.002)

f_inc	0.010	0.014
	(0.032)	(0.032)
eqinc	0.117	0.084
	(0.294)	(0.294)
r_d	0.015	0.019
	(0.030)	(0.030)
r_d_blank	-0.001	-0.002
	(0.007)	(0.007)
HHI	-0.00001**	-0.00001**
	(0.00000)	(0.00000)
Control variables	Yes	Yes
Individual fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Shock	5% - 15%	15% ->
Observations	79,506	79,506
R^2	0.006	0.006
Significance levels		*p<0.1; **p<0.05; ***p<0.01

Panel b: Cash ETR

Event: Shock in cash ETR

	Dependent variable: Blockholder's portfolio weight	
	(1)	(2)
5+ years before neg shock	0.006*	-0.003
	(0.004)	(0.005)
4 years before neg shock	-0.001	-0.004
	(0.002)	(0.003)
3 years before neg shock	0.001	-0.005
	(0.002)	(0.003)
2 years before neg shock	-0.001	-0.003
	(0.002)	(0.002)
Negative shock year	-0.001	0.001
	(0.001)	(0.002)
1 year after neg shock	-0.004**	-0.003
	(0.002)	(0.003)
2 years after neg shock	-0.005**	0.004
	(0.002)	(0.003)
3 years after neg shock	-0.009***	0.005

	(0.002)	(0.004)
4 years after neg shock	-0.008***	0.002
	(0.003)	(0.004)
5+ years after neg shock	-0.007	-0.002
	(0.004)	(0.005)
5+ years before pos shock	-0.0001	-0.003
	(0.004)	(0.004)
4 years before pos shock	-0.003	-0.004
	(0.002)	(0.003)
3 years before pos shock	-0.003	0.003
	(0.002)	(0.003)
2 years before pos shock	-0.002	0.002
	(0.002)	(0.002)
Positive shock year	-0.002	0.002
	(0.001)	(0.002)
1 year after pos shock	0.001	-0.002
	(0.002)	(0.002)
2 years after pos shock	-0.002	0.002
	(0.002)	(0.003)
3 years after pos shock	-0.002	-0.001
	(0.002)	(0.003)
4 years after pos shock	-0.002	-0.007**
	(0.002)	(0.003)
5+ years after pos shock	-0.0004	-0.001
	(0.004)	(0.005)
intan	-0.021**	-0.022**
	(0.009)	(0.009)
inst	-0.006	-0.007
	(0.007)	(0.007)
lev	0.017**	0.017**
	(0.008)	(0.008)
mb	0.0002	0.0002
	(0.0002)	(0.0002)
nol	-0.002	-0.002
	(0.002)	(0.002)
diffnol	-0.025	-0.028
	(0.027)	(0.027)
ppe	-0.016	-0.017

	(0.016)	(0.016)
roa	-0.002	0.002
	(0.015)	(0.016)
log_mktcap	0.019***	0.019***
	(0.002)	(0.002)
f_inc	0.011	0.013
	(0.031)	(0.032)
eqinc	0.119	0.116
	(0.294)	(0.295)
r_d	0.016	0.015
	(0.030)	(0.030)
r_d_blank	-0.001	-0.001
	(0.007)	(0.007)
HHI	-0.00001**	-0.00001**
	(0.00000)	(0.00000)
Control variables	Yes	Yes
Individual fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Shock	5% - 15%	15% ->
Observations	79,506	79,506
R^2	0.006	0.006
Significance levels		*p<0.1; **p<0.05; ***p<0.01

Appendixes

Appendix A: Variable definitions

Measures of Tax Avoidance

GAAP ETR Effective tax rate (Compustat: TXT/PI). ϵ [0,1]

CASH ETR Cash effective tax rate (Compustat: TXPD/PI), ϵ [0,1]

Variables of interest

Peer ETR Based on GAAP ETR (CASH ETR) when using GAAP ETR (CASH ETR) as a tax avoidance measure

Control Variables

ROA Return on assets defined as pretax income scaled by total assets. (Compustat: PI/AT)

INTAN	Intangibles assets: computed as reported intangibles (Compustat: INTAN) divided by total assets (Compustat: AT). Missing values of INTAN is set equal to 0
INST	Institutional ownership %: Calculated in three stages. First shares (ownership data) / (Compustat: CSHOQ) to get institutional ownership of every investor for each quarter. In the next stage we sum the institutional ownership per firm each quarter. And in the last stage we take the mean/collapse the quarters to yearly values.
LEV	Leverage: computed as long-term debt (Compustat: DLTT) divided by total assets (Compustat: AT). Missing values of LEV is set equal to 0
MB	Market-to-book ratio: computed as the ratio of market value of equity (Compustat: PRCC_F*CSHO) to book value of equity (Compustat: CEQ). Missing values of MB is set to 0
NOL	Dummy equal to 1 if reported positive tax loss carryforward during the year (Compustat: TLCF) and 0 for rest
DIFFNOL	Change in NOL: computed as the change in firm i's NOL (Compustat: TLCF) between year t and year t-1, scaled by total assets (Compustat: AT)
PPE	Property, plant and equipment: computed as net property, plant and equipment (Compustat: PPENT), divided by total assets (Compustat: AT). Missing values of PPE is set equal to 0
LOG_MKTCAP	Log of total market capital (Compustat: PRCC_F*CSHO). Missing values is set to the mean across the sample
F_INC	Foreign pretax income (Compustat: PIFO) scaled by total assets (Compustat: AT). Missing values of F_INC is set to 0
R_D	Research & Development: computed as R&D expenses (Compustat: XRD) scaled by total assets (Compustat: AT). Missing variables of R_D is set either industry average (2-digit SIC) where available or equal to 0
R_D_Blank	Dummy equal to 1 for missing values and 0 for non-missing values of R&D
EQINC	Equity method earnings: computed as equity in earnings (Compustat: ESUB) divided by total assets (Compustat: AT). Missing observations of EQINC is set equal to 0
HHI	Herfindahl-Hirschman Index calculated based on 2-digit SIC

Appendix B: Descriptive statistics of number of shocks in the sample from event study on peer ETR's

Panel A: GAAP ETR

Descriptive statistics for Event study		
Interval →	-5% >= SHOCK > -10%	5% <= SHOCK < 10%
4 shocks for the same firm	0	1
3 shocks for the same firm	7	6
2 shocks for the same firm	75	49
1 shock for the same firm	507	420
SUM shocks	678	540
Interval →	-10% >= SHOCK	10% <= SHOCK
3 shocks for the same firm	0	4
2 shocks for the same firm	9	11
1 shock for the same firm	198	197
SUM shocks	216	231

Panel B: Cash ETR

Descriptive statistics for Event study		
Interval →	-5% >= SHOCK > -10%	5% <= SHOCK < 10%
5 shocks for the same firm	0	2
4 shocks for the same firm	6	8
3 shocks for the same firm	51	51
2 shocks for the same firm	196	251
1 shock for the same firm	802	834
SUM shocks	1371	1531
Interval →	-10% >= SHOCK	10% <= SHOCK
5 shocks for the same firm	0	1
4 shocks for the same firm	0	1
3 shocks for the same firm	9	6
2 shocks for the same firm	49	52
1 shock for the same firm	423	437
SUM shocks	548	568

Appendix C: Missing values handling

Panel A: GAAP ETR sample

Observations before introducing lag model	23,603
Loss when introducing lag model	(6,600)

Net firm-year observations in the sample	17,003	
Missing values for control variables in net sample		
Variable name	**Missing observations**	**Handling**
intan (Intangible assets deflated by total assets)	441	Set to 0.
lev (Long term debt deflated by total assets)	42	Set to 0.
mb (Market value of equity deflated by book value of equity)	5	Set to 0.
ppe (Property, plant and equipment deflated by total assets)	8	Set to 0.
roa (Pre-tax income deflated by total assets)	0	-
log_mktcap (Natural log of market capitalization / market value of equity)	4	Set to mean
f_inc (Foreign pre-tax income deflated by total assets)	7,209	This number is after replacing missing pifo with pi-pidom. Rest is set to 0.
eqinc (Equity in earnings deflated by total assets)	1,873	Set to 0.
r_d (Research and development expense deflated by total assets)	608	This number is after replacing missing values with industry average. Rest is set to 0.
r_d_blank (dummy for missing values in research and development before replacement with industry average)	0	-
HHI (Herfindahl-Hirschman Index)	0	-
inst (Institutional ownership %)	0	-
tlcf (Tax loss carry forward)	4,556	Set to 0.
diffnol (Difference in tax loss carry forward between year t and t-1)	0	This is after setting tlcf to 0.
nol (Dummy for positive tlcf)	0	This is after setting tlcf to 0.

Panel B: Cash ETR sample

Observations before introducing lag model	23,015	
Loss when introducing lag model	(6,377)	
Net firm-year observations in the sample	16,638	
Missing values for control variables in net sample		
Variable name	**Missing observations**	**Handling**
intan (Intangible assets deflated by total assets)	426	Set to 0.
lev (Long term debt deflated by total assets)	53	Set to 0.
mb (Market value of equity deflated by book value of equity)	1	Set to 0.
ppe (Property, plant and equipment deflated by total assets)	7	Set to 0.
roa (Pre-tax income deflated by total assets)	0	-
log_mktcap (Natural log of market capitalization / market value of equity)	0	-
f_inc (Foreign pre-tax income deflated by total assets)	6,842	This number is after replacing missing pifo with pi-pidom. Rest is set to 0.
eqinc (Equity in earnings deflated by total assets)	1,788	Set to 0.
r_d (Research and development expense deflated by total assets)	579	This number is after replacing missing values with industry average. Rest is set to 0.
r_d_blank (dummy for missing values in research and development before replacement with industry average)	0	-
HHI (Herfindahl-Hirschman Index)	0	-
inst (Institutional ownership %)	0	-
tlcf (Tax loss carry forward)	4,429	Set to 0.

diffnol (Difference in tax loss carry forward between year t and t-1)	0	This is after setting tlcf to 0.
nol (Dummy for positive tlcf)	0	This is after setting tlcf to 0.